

WO02083953

Publication Title:

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

Abstract:

The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

Data supplied from the esp@cenet database - <http://ep.espacenet.com>

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
24 October 2002 (24.10.2002)

PCT

(10) International Publication Number
WO 02/083953 A1

(51) International Patent Classification⁷: C12Q 1/68, C07H 21/02, G01N 27/26

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(21) International Application Number: PCT/US02/11757

(84) Designated States (*regional*): ARIGO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(22) International Filing Date: 11 April 2002 (11.04.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/282,965 11 April 2001 (11.04.2001) US

Published:

- with international search report
- with amended claims

(71) Applicant (*for all designated States except US*): PTC THERAPEUTICS, INC. [US/US]; 100 Corporate Court, Middlesex Business Center, South Plainfield, NJ 07080 (US).

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): RANDO, Robert [US/US]; 3 Brown Court, Annandale, NJ 08801 (US). WELCH, Ellen [US/US]; 33 Hollow Brook Road, Califon, NJ 07830 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

WO 02/083953 A1

(54) Title: METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

(57) Abstract: The present invention relates to a method for screening and identifying test compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-competitive binding assays are advantageously used to screen libraries of compounds for those that selectively bind to a preselected target RNA. Binding of target RNA molecules to a particular test compound is detected using any physical method that measures the altered physical property of the target RNA bound to a test compound. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. The methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of compounds to identify pharmaceutical leads.

METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

5 This application claims the benefit of U.S. Provisional Application No.
60/282,965, filed April 11, 2001, which is incorporated herein by reference in its entirety.

1. INTRODUCTION

The present invention relates to a method for screening and identifying test
10 compounds that bind to a preselected target ribonucleic acid ("RNA"). Direct, non-
competitive binding assays are advantageously used to screen libraries of compounds for
those that selectively bind to a preselected target RNA. Binding of target RNA molecules to
a particular test compound is detected using any physical method that measures the altered
physical property of the target RNA bound to a test compound. The methods of the present
15 invention provide a simple, sensitive assay for high-throughput screening of libraries of
compounds to identify pharmaceutical leads.

2. BACKGROUND OF THE INVENTION

Protein-nucleic acid interactions are involved in many cellular functions,
20 including transcription, RNA splicing, mRNA decay, and mRNA translation. Readily
accessible synthetic molecules that can bind with high affinity to specific sequences of
single- or double-stranded nucleic acids have the potential to interfere with these
interactions in a controllable way, making them attractive tools for molecular biology and
medicine. Successful approaches for blocking function of target nucleic acids include using
25 duplex-forming antisense oligonucleotides (Miller, 1996, Progress in Nucl. Acid Res. &
Mol. Biol. 52:261-291; Ojwang & Rando, 1999, Achieving antisense inhibition by
oligodeoxynucleotides containing N, modified 2'-deoxyguanosine using tumor necrosis
factor receptor type 1, METHODS: A Companion to Methods in Enzymology 18:244-251)
and peptide nucleic acids ("PNA") (Nielsen, 1999, Current Opinion in Biotechnology
30 10:71-75), which bind to nucleic acids via Watson-Crick base-pairing. Triplex-forming
anti-gene oligonucleotides can also be designed (Ping *et al.*, 1997, RNA 3:850-860;
Aggarwal *et al.*, 1996, Cancer Res. 56:5156-5164; U.S. Patent No. 5,650,316), as well as
pyrrole-imidazole polyamide oligomers (Gottesfeld *et al.*, 1997, Nature 387:202-205; White
et al., 1998, Nature 391:468-471), which are specific for the major and minor grooves of a
35 double helix, respectively.

In addition to synthetic nucleic acids (*i.e.*, antisense, ribozymes, and triplex-forming molecules), there are examples of natural products that interfere with deoxyribonucleic acid ("DNA") or RNA processes such as transcription or translation. For example, certain carbohydrate-based host cell factors, calicheamicin oligosaccharides, 5 interfere with the sequence-specific binding of transcription factors to DNA and inhibit transcription *in vivo* (Ho *et al.*, 1994, Proc. Natl. Acad. Sci. USA 91:9203-9207; Liu *et al.*, 1996, Proc. Natl. Acad. Sci. USA 93:940-944). Certain classes of known antibiotics have been characterized and were found to interact with RNA. For example, the antibiotic 10 thiostreptone binds tightly to a 60-mer from ribosomal RNA (Cundliffe *et al.*, 1990, in *The Ribosome: Structure, Function & Evolution* (Schlessinger *et al.*, eds.) American Society for Microbiology, Washington, D.C. pp. 479-490). Bacterial resistance to various antibiotics often involves methylation at specific rRNA sites (Cundliffe, 1989, Ann. Rev. Microbiol. 43:207-233). Aminoglycosidic aminocyclitol (aminoglycoside) antibiotics and peptide 15 antibiotics are known to inhibit group I intron splicing by binding to specific regions of the RNA (von Ahsen *et al.*, 1991, Nature (London) 353:368-370). Some of these same aminoglycosides have also been found to inhibit hammerhead ribozyme function (Stage *et al.*, 1995, RNA 1:95-101). In addition, certain aminoglycosides and other protein synthesis 20 inhibitors have been found to interact with specific bases in 16S rRNA (Woodcock *et al.*, 1991, EMBO J. 10:3099-3103). An oligonucleotide analog of the 16S rRNA has also been shown to interact with certain aminoglycosides (Purohit *et al.*, 1994, Nature 370:659-662). A molecular basis for hypersensitivity to aminoglycosides has been found to be located in a single base change in mitochondrial rRNA (Hutchin *et al.*, 1993, Nucleic Acids Res. 25 21:4174-4179). Aminoglycosides have also been shown to inhibit the interaction between specific structural RNA motifs and the corresponding RNA binding protein. Zapp *et al.* (Cell, 1993, 74:969-978) has demonstrated that the aminoglycosides neomycin B, 30 lividomycin A, and tobramycin can block the binding of Rev, a viral regulatory protein required for viral gene expression, to its viral recognition element in the IIB (or RRE) region of HIV RNA. This blockage appears to be the result of competitive binding of the antibiotics directly to the RRE RNA structural motif.

Single stranded sections of RNA can fold into complex tertiary structures consisting of local motifs such as loops, bulges, pseudoknots, guanosine quartets and turns (Chastain & Tinoco, 1991, Progress in Nucleic Acid Res. & Mol. Biol. 41:131-177; Chow & Bogdan, 1997, Chemical Reviews 97:1489-1514; Rando & Hogan, 1998, Biologic 35 activity of guanosine quartet forming oligonucleotides in "Applied Antisense Oligonucleotide Technology" Stein. & Krieg (eds) John Wiley and Sons, New York, pages

335-352). Such structures can be critical to the activity of the nucleic acid and affect functions such as regulation of mRNA transcription, stability, or translation (Weeks & Crothers, 1993; Science 261:1574-1577). The dependence of these functions on the native three-dimensional structural motifs of single-stranded stretches of nucleic acids makes it difficult to identify or design synthetic agents that bind to these motifs using general, simple-to-use sequence-specific recognition rules for the formation of double- and triple-helical nucleic acids used in the design of antisense and ribozyme type molecules.

Approaches to screening generally involve competitive assays designed to identify compounds that disrupt the interaction between a target RNA and a physiological, host cell factor(s) that had been previously identified to specifically interact with that particular target RNA. In general, such assays require the identification and characterization of the host cell factor(s) deemed to be required for the function of the target RNA. Both the target RNA and its preselected host cell binding partner are used in a competitive format to identify compounds that disrupt or interfere with the two components in the assay.

Citation or identification of any reference in Section 2 of this application is not an admission that such reference is available as prior art to the present invention.

3. SUMMARY OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids including, but not limited to, specific RNA sequences, RNA structural motifs, and/or RNA structural elements. The specific target RNA sequences, RNA structural motifs, and/or RNA structural elements are used as targets for screening small molecules and identifying those that directly bind these specific sequences, motifs, and/or structural elements. For example, methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds, preferably under physiologic conditions. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the uncomplexed, unlabeled test compounds in the library, and from uncomplexed, labeled target RNA, by a variety of methods, including but not limited to, methods that differentiate changes in the electrophoretic, chromatographic, or thermostable

properties of the complexed target RNA. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. The structure of the test compound attached to the labeled RNA is then determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds, each having an address or identifier, may be deconvoluted, *e.g.*, by cross-referencing the positive sample to original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR"). The test compounds identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and the like. In addition, small organic molecules which interact specifically with target RNA molecules may be useful as lead compounds for the development of therapeutic agents.

The methods described herein for the identification of compounds that directly bind to a particular preselected target RNA are well suited for high-throughput screening. The direct binding method of the invention offers advantages over drug screening systems for competitors that inhibit the formation of naturally-occurring RNA binding protein:target RNA complexes; *i.e.*, competitive assays. The direct binding method of the invention is rapid and can be set up to be readily performed, *e.g.*, by a technician, making it amenable to high throughput screening. The method of the invention also eliminates the bias inherent in the competitive drug screening systems, which require the use of a preselected host cell factor that may not have physiological relevance to the activity of the target RNA. Instead, the methods of the invention are used to identify any compound that can directly bind to specific target RNA sequences, RNA structural motifs, and/or RNA structural elements, preferably under physiologic conditions. As a result, the compounds so identified can inhibit the interaction of the target RNA with any one or more of the native host cell factors (whether known or unknown) required for activity of the RNA *in vivo*.

The present invention may be understood more fully by reference to the detailed description and examples, which are intended to illustrate non-limiting embodiments of the invention.

3.1. Definitions

As used herein, a "target nucleic acid" refers to RNA, DNA, or a chemically modified variant thereof. In a preferred embodiment, the target nucleic acid is RNA. A target nucleic acid also refers to tertiary structures of the nucleic acids, such as, but not limited to loops, bulges, pseudoknots, guanosine quartets and turns. A target nucleic acid also refers to RNA elements such as, but not limited to, the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, which are described in Section 5.1. Non-limiting examples of target nucleic acids are presented in Section 5.1 and Section 6.

As used herein, a "library" refers to a plurality of test compounds with which a target nucleic acid molecule is contacted. A library can be a combinatorial library, *e.g.*, a collection of test compounds synthesized using combinatorial chemistry techniques, or a collection of unique chemicals of low molecular weight (less than 1000 daltons) that each occupy a unique three-dimensional space.

As used herein, a "label" or "detectable label" is a composition that is detectable, either directly or indirectly, by spectroscopic, photochemical, biochemical, immunochemical, or chemical means. For example, useful labels include radioactive isotopes (*e.g.*, ^{32}P , ^{35}S , and ^3H), dyes, fluorescent dyes, electron-dense reagents, enzymes and their substrates (*e.g.*, as commonly used in enzyme-linked immunoassays, *e.g.*, alkaline phosphatase and horse radish peroxidase), biotin-streptavidin, digoxigenin, or haptens and proteins for which antisera or monoclonal antibodies are available. Moreover, a label or detectable moiety can include a "affinity tag" that, when coupled with the target nucleic acid and incubated with a test compound or compound library, allows for the affinity capture of the target nucleic acid along with molecules bound to the target nucleic acid. One skilled in the art will appreciate that a affinity tag bound to the target nucleic acids has, by definition, a complimentary ligand coupled to a solid support that allows for its capture. For example, useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamers, or haptens and proteins for which antisera or monoclonal antibodies are available. The label or detectable moiety is typically bound, either covalently, through a linker or chemical bound, or through ionic, van der Waals or hydrogen bonds to the molecule to be detected.

As used herein, a "dye" refers to a molecule that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means.

As used herein, a "visible dye" refers to a molecule having a chromophore that absorbs radiation in the visible region of the spectrum (*i.e.*, having a wavelength of between about 400 nm and about 700 nm) such that the transmitted radiation is in the visible region and can be detected either visually or by conventional spectroscopic means. As used herein, an "ultraviolet dye" refers to a molecule having a chromophore that absorbs radiation in the ultraviolet region of the spectrum (*i.e.*, having a wavelength of between about 30 nm and about 400 nm). As used herein, an "infrared dye" refers to a molecule having a chromophore that absorbs radiation in the infrared region of the spectrum (*i.e.*, having a wavelength between about 700 nm and about 3,000 nm). A "chromophore" is the network of atoms of the dye that, when exposed to radiation, emits radiation at a level that is detectable visually or via conventional spectroscopic means. One of skill in the art will readily appreciate that although a dye absorbs radiation in one region of the spectrum, it may emit radiation in another region of the spectrum. For example, an ultraviolet dye may emit radiation in the visible region of the spectrum. One of skill in the art will also readily appreciate that a dye can transmit radiation or can emit radiation via fluorescence or phosphorescence.

The phrase "pharmaceutically acceptable salt(s)," as used herein includes but is not limited to salts of acidic or basic groups that may be present in test compounds identified using the methods of the present invention. Test compounds that are basic in nature are capable of forming a wide variety of salts with various inorganic and organic acids. The acids that can be used to prepare pharmaceutically acceptable acid addition salts of such basic compounds are those that form non-toxic acid addition salts, *i.e.*, salts containing pharmacologically acceptable anions, including but not limited to sulfuric, citric, maleic, acetic, oxalic, hydrochloride, hydrobromide, hydroiodide, nitrate, sulfate, bisulfate, phosphate, acid phosphate, isonicotinate, acetate, lactate, salicylate, citrate, acid citrate, tartrate, oleate, tannate, pantothenate, bitartrate, ascorbate, succinate, maleate, gentisinate, fumarate, gluconate, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate and pamoate (*i.e.*, 1,1'-methylene-bis-(2-hydroxy-3-naphthoate)) salts. Test compounds that include an amino moiety may form pharmaceutically or cosmetically acceptable salts with various amino acids, in addition to the acids mentioned above. Test compounds that are acidic in nature are capable of forming base salts with various pharmacologically or cosmetically acceptable cations. Examples of such salts include alkali metal or alkaline earth metal salts and, particularly, calcium, magnesium, sodium lithium, zinc, potassium, and iron salts.

By "substantially one type of test compound," as used herein, is meant that the assay can be performed in such a fashion that at some point, only one compound need be used in each reaction so that, if the result is indicative of a binding event occurring between the target RNA molecule and the test compound, the test compound can be easily identified.

5

4. DESCRIPTION OF DRAWINGS

- FIG. 1. Gel retardation analysis to detect peptide-RNA interactions. In 20 μ l reactions containing increasing concentrations of Tat₄₇₋₅₈ peptide (0.1 μ M, 0.2 μ M, 0.4 μ M, 0.8 μ M, 1.6 μ M) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90 °C for 2 min and allowed to cool slowly to 24 °C. 10 ml of 30% glycerol was added to each sample and applied to a 12% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4 °C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated with a phosphorimager. The concentration of peptide added is indicated above each lane.
- 10
15
- FIG. 2. Gentamicin interacts with an oligonucleotide corresponding to the 16S rRNA. 20 μ l reactions containing increasing concentrations of gentamicin (1 ng/ml, 10 ng/ml, 100 ng/ml, 1 μ g/ml, 10 μ g/ml, 50 μ g/ml, 500 μ g/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer, heated at 90 °C for 2 min and allowed to cool slowly to 24 °C. Then 10 μ l of 30% glycerol was added to each sample and the samples were applied to a 13.5% non-denaturing polyacrylamide gel. The gel was electrophoresed using 1200 volt-hours at 4 °C in TBE Buffer. Following electrophoresis, the gel was dried and the radioactivity was quantitated using a phosphorimager. The concentration of gentamicin added is indicated above each lane.
- 20
25
- FIG. 3. The presence of 10 pg/ml gentamicin produces a gel mobility shift in the presence of the 16S rRNA oligonucleotide. 20 μ l reactions containing increasing concentrations of gentamicin (100 ng/ml, 10 ng/ml, 1 ng/ml, 100 pg/ml, and 10 pg/ml) were added to 50 pmole RNA oligonucleotide in TKM buffer were treated as described for Figure 2.
- 30
- FIG. 4. Gentamicin binding to the 16S rRNA oligonucleotide is weak in the absence of MgCl₂. Reaction mixtures containing gentamicin (1 mg/ml, 100 μ g/ml,
- 35

10 µg/ml, 1 µg/ml, 0.1 µg/ml, and 10 ng/ml) were treated as described in Figure 2 except that the TKM buffer does not contain MgCl₂.

FIG. 5. Gel retardation analysis to detect peptide-RNA interactions. In reactions containing increasing concentrations of Tat₄₇₋₅₈ peptide (0.1 µM, 0.2 µM, 0.4 µM, 0.8 µM, 1.6 µM) 50 pmole TAR RNA oligonucleotide was added in TK buffer. The reaction mixture was then heated at 90°C for 2 min and allowed to cool slowly to 24°C. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania). The peaks correspond to the amount of free TAR RNA ("TAR") or the Tat-TAR complex ("Tat-TAR"). The concentration of peptide added is indicated below each lane.

5. DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to methods for identifying compounds that bind to preselected target elements of nucleic acids, in particular, RNAs, including but not limited to preselected target RNA sequencing structural motifs, or structural elements. Methods are described in which a preselected target RNA having a detectable label is used to screen a library of test compounds. Any complexes formed between the target RNA and a member of the library are identified using physical methods that detect the altered physical property of the target RNA bound to a test compound. Changes in the physical property of the RNA-test compound complex relative to the target RNA or test compound can be measured by methods such as, but not limited to, methods that detect a change in mobility due to a change in mass, change in charge, or a change in thermostability. Such methods include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation. In particular, the present invention relates to methods for using a target RNA having a detectable label to screen a library of test compounds free in solution, in labeled tubes or microtiter plate, or in a microarray. Compounds in the library that bind to the labeled target RNA will form a detectably labeled complex. The detectably labeled complex can then be identified and removed from the unlabeled, uncomplexed test compounds in the library by a variety of methods capable of differentiating changes in the physical properties of the complexed target RNA. The structure of the test compound attached to the labeled RNA is also determined. The methods used will depend, in part, on the nature of the library screened. For example, assays or microarrays of test compounds,

each having an address or identifier, may be deconvoluted, e.g., by cross-referencing the positive sample to an original compound list that was applied to the individual test assays. Another method for identifying test compounds includes *de novo* structure determination of the test compounds using mass spectrometry or nuclear magnetic resonance ("NMR").

5 Thus, the methods of the present invention provide a simple, sensitive assay for high-throughput screening of libraries of test compounds, in which the test compounds of the library that specifically bind a preselected target nucleic acid are easily distinguished from non-binding members of the library. The structures of the binding molecules are deciphered from the input library by methods depending on the type of library that is used.

10 The test compounds so identified are useful for any purpose to which a binding reaction may be put, for example in assay methods, diagnostic procedures, cell sorting, as inhibitors of target molecule function, as probes, as sequestering agents and lead compounds for development of therapeutics, and the like. Small organic compounds that are identified to

15 interact specifically with the target RNA molecules are particularly attractive candidates as lead compounds for the development of therapeutic agents.

The assay of the invention reduces bias introduced by competitive binding assays which require the identification and use of a host cell factor (presumably essential for modulating RNA function) as a binding partner for the target RNA. The assays of the 20 present invention are designed to detect any compound or agent that binds to the target RNA, preferably under physiologic conditions. Such agents can then be tested for biological activity, without establishing or guessing which host cell factor or factors is required for modulating the function and/or activity of the target RNA.

Section 5.1 describes examples of protein-RNA interactions that are 25 important in a variety of cellular functions and several target RNA elements that can be used to identify test compounds. Compounds that inhibit these interactions by binding to the RNA and successfully competing with the natural protein or host cell factor that endogenously binds to the RNA may be important, e.g., in treating or preventing a disease or abnormal condition, such as an infection or unchecked growth. Section 5.2 describes detectable labels for target nucleic acids that are useful in the methods of the invention.

30 Section 5.3 describes libraries of test compounds. Section 5.4 provides conditions for binding a labeled target RNA to a test compound of a library and detecting RNA binding to a test compound using the methods of the invention. Section 5.5 provides methods for separating complexes of target RNAs bound to a test compound from an unbound RNA.

35 Section 5.6 describes methods for identifying test compounds that are bound to the target RNA. Section 5.7 describes a secondary, biological screen of test compounds identified by

the methods of the invention to test the effect of the test compounds *in vivo*. Section 5.8 describes the use of test compounds identified by the methods of the invention for treating or preventing a disease or abnormal condition in mammals.

5

5.1. Biologically Important RNA-Host Cell Factor Interactions

Nucleic acids, and in particular RNAs, are capable of folding into complex tertiary structures that include bulges, loops, triple helices and pseudoknots, which can provide binding sites for host cell factors, such as proteins and other RNAs. RNA-protein and RNA-RNA interactions are important in a variety cellular functions, including transcription, RNA splicing, RNA stability and translation. Furthermore, the binding of such host cell factors to RNAs may alter the stability and translational efficiency of such RNAs, and according affect subsequent translation. For example, some diseases are associated with protein overproduction or decreased protein function. In this case, the identification of compounds to modulate RNA stability and translational efficiency will be useful to treat and prevent such diseases.

The methods of the present invention are useful for identifying test compounds that bind to target RNA elements in a high throughput screening assay of libraries of test compounds in solution. In particular, the methods of the present invention are useful for identifying a test compound that binds to a target RNA elements and inhibits the interaction of that RNA with one or more host cell factors *in vivo*. The molecules identified using the methods of the invention are useful for inhibiting the formation of a specific bound RNA:host cell factor complexes *in vivo*.

In some embodiments, test compounds identified by the methods of the invention are useful for increasing or decreasing the translation of messenger RNAs (“mRNAs”), e.g., protein production, by binding to one or more regulatory elements in the 5' untranslated region, the 3' untranslated region, or the coding region of the mRNA. Compounds that bind to mRNA can, *inter alia*, increase or decrease the rate of mRNA processing, alter its transport through the cell, prevent or enhance binding of the mRNA to ribosomes, suppressor proteins or enhancer proteins, or alter mRNA stability. Accordingly, compounds that increase or decrease mRNA translation can be used to treat or prevent disease. For example, diseases associated with protein overproduction, such as amyloidosis, or with the production of mutant proteins, such as *Ras*, can be treated or prevented by decreasing translation of the mRNA that codes for the overproduced protein, thus inhibiting production of the protein. Conversely, the symptoms of diseases associated with decreased protein function, such as hemophilia, may be treated by increasing

translation of mRNA coding for the protein whose function is decreased, e.g., factor IX in some forms of hemophilia.

The methods of the invention can be used to identify compounds that bind to mRNAs coding for a variety of proteins with which the progression of diseases in mammals is associated. These mRNAs include, but are not limited to, those coding for amyloid protein and amyloid precursor protein; anti-angiogenic proteins such as angiostatin, endostatin, METH-1 and METH-2; apoptosis inhibitor proteins such as survivin, clotting factors such as Factor IX, Factor VIII, and others in the clotting cascade; collagens; cyclins and cyclin inhibitors, such as cyclin dependent kinases, cyclin D1, cyclin E, WAF1, cdk4 inhibitor, and MTS1; cystic fibrosis transmembrane conductance regulator gene (CFTR); cytokines such as IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL-10, IL-11, IL-12, IL-13, IL-14, IL-15, IL-16, IL-17 and other interleukins; hematopoietic growth factors such as erythropoietin (Epo); colony stimulating factors such as G-CSF, GM-CSF, M-CSF, SCF and thrombopoietin; growth factors such as BDNF, BMP, GGRP, EGF, FGF, GDNF, GGF, HGF, IGF-1, IGF-2, KGF, myotrophin, NGF, OSM, PDGF, somatotrophin, TGF- β , TGF- α and VEGF; antiviral cytokines such as interferons, antiviral proteins induced by interferons, TNF- α , and TNF- β ; enzymes such as cathepsin K, cytochrome P-450 and other cytochromes, farnesyl transferase, glutathione-S transferases, heparanase, HMG CoA synthetase, N-acetyltransferase, phenylalanine hydroxylase, phosphodiesterase, ras carboxyl-terminal protease, telomerase and TNF converting enzyme; glycoproteins such as cadherins, e.g., N-cadherin and E-cadherin; cell adhesion molecules; selectins; transmembrane glycoproteins such as CD40; heat shock proteins; hormones such as 5- α reductase, atrial natriuretic factor, calcitonin, corticotrophin releasing factor, diuretic hormones, glucagon, gonadotropin, gonadotropin releasing hormone, growth hormone, growth hormone releasing factor, somatotropin, insulin, leptin, luteinizing hormone, luteinizing hormone releasing hormone, parathyroid hormone, thyroid hormone, and thyroid stimulating hormone; proteins involved in immune responses, including antibodies, CTLA4, hemagglutinin, MHC proteins, VLA-4, and kallikrein-kininogen-kinin system; ligands such as CD4; oncogene products such as *sis*, *hst*, protein tyrosine kinase receptors, *ras*, *abl*, *mos*, *myc*, *fos*, *jun*, *H-ras*, *ki-ras*, *c-fms*, *bcl-2*, *L-myc*, *c-myc*, *gip*, *gsp*, and *HER-2*; receptors such as bombesin receptor, estrogen receptor, GABA receptors, growth factor receptors including EGFR, PDGFR, FGFR, and NGFR, GTP-binding regulatory proteins, interleukin receptors, ion channel receptors, leukotriene receptor antagonists, lipoprotein receptors, opioid pain receptors, substance P receptors, retinoic acid and retinoid receptors, steroid receptors, T-cell receptors, thyroid hormone receptors, TNF receptors; tissue

plasminogen activator; transmembrane receptors; transmembrane transporting systems, such as calcium pump, proton pump, Na/Ca exchanger, MRP1, MRP2, P170, LRP, and cMOAT; transferrin; and tumor suppressor gene products such as *APC*, *brcal*, *brca2*, *DCC*, *MCC*, *MTS1*, *NF1*, *NF2*, *nm23*, *p53* and *Rb*. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic. Other target genes include, but are not limited to, those disclosed in Section 5.1 and Section 6.

The methods of the invention can be used to identify mRNA-binding test compounds for increasing or decreasing the production of a protein, thus treating or preventing a disease associated with decreasing or increasing the production of said protein, respectively. The methods of the invention may be useful for identifying test compounds for treating or preventing a disease in mammals, including cats, dogs, swine, horses, goats, sheep, cattle, primates and humans. Such diseases include, but are not limited to, amyloidosis, hemophilia, Alzheimer's disease, atherosclerosis, cancer, giantism, dwarfism, hypothyroidism, hyperthyroidism, inflammation, cystic fibrosis, autoimmune disorders, diabetes, aging, obesity, neurodegenerative disorders, and Parkinson's disease. Other diseases include, but are not limited to, those described in Section 5.1 and diseases caused by aberrant expression of the genes disclosed in Example 6. In addition to the eukaryotic genes listed above, the invention, as described, can be used to define molecules that interrupt viral, bacterial or fungal transcription or translation efficiencies and therefore form the basis for a novel anti-infectious disease therapeutic.

In other embodiments, test compounds identified by the methods of the invention are useful for preventing the interaction of an RNA, such as a transfer RNA ("tRNA"), an enzymatic RNA or a ribosomal RNA ("rRNA"), with a protein or with another RNA, thus preventing, e.g., assembly of an *in vivo* protein-RNA or RNA-RNA complex that is essential for the viability of a cell. The term "enzymatic RNA," as used herein, refers to RNA molecules that are either self-splicing, or that form an enzyme by virtue of their association with one or more proteins, e.g., as in RNase P, telomerase or small nuclear ribonuclear protein particles. For example, inhibition of an interaction between rRNA and one or more ribosomal proteins may inhibit the assembly of ribosomes, rendering a cell incapable of synthesizing proteins. In addition, inhibition of the interaction of precursor rRNA with ribonucleases or ribonucleoprotein complexes (such as RNase P) that process the precursor rRNA prevent maturation of the rRNA and its assembly into ribosomes. Similarly, a tRNA:tRNA synthetase complex may be inhibited by test

compounds identified by the methods of the invention such that tRNA molecules do not become charged with amino acids. Such interactions include, but are not limited to, rRNA interactions with ribosomal proteins, tRNA interactions with tRNA synthetase, RNase P protein interactions with RNase P RNA, and telomerase protein interactions with telomerase RNA.

In other embodiments, test compounds identified by the methods of the invention are useful for treating or preventing a viral, bacterial, protozoan or fungal infection. For example, transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation response region RNA ("TAR RNA"). HIV TAR RNA is a 59-base stem-loop structure located at the 5'-end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a potential binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 upregulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA are useful as anti-HIV therapeutics (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Hamy *et al.*, 1998, *Biochemistry* 37:5086-5095; Mei *et al.*, 1998, *Biochemistry* 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

The methods of the invention can be used to identify test compounds to treat or prevent viral, bacterial, protozoan or fungal infections in a patient. In some embodiments, the methods of the invention are useful for identifying compounds that decrease translation of microbial genes by interacting with mRNA, as described above, or for identifying compounds that inhibit the interactions of microbial RNAs with proteins or other ligands that are essential for viability of the virus or microbe. Examples of microbial target RNAs useful in the present invention for identifying antiviral, antibacterial, anti-protozoan and anti-fungal compounds include, but are not limited to, general antiviral and anti-inflammatory targets such as mRNAs of INF α , INF γ , RNase L, RNase L inhibitor protein, PKR, tumor necrosis factor, interleukins 1-15, and IMP dehydrogenase; internal ribosome entry sites; HIV-1 CT rich domain and RNase H mRNA; HCV internal ribosome entry site (required to direct translation of HCV mRNA), and the 3'-untranslated tail of HCV genomes; rotavirus NSP3 binding site, which binds the protein NSP3 that is required for rotavirus mRNA translation; HBV epsilon domain; Dengue virus 5' and 3' untranslated regions, including IRES; INF α , INF β and INF γ ; plasmodium falciparum mRNAs; the 16S

ribosomal subunit ribosomal RNA and the RNA component of RNase P of bacteria; and the RNA component of telomerase in fungi and cancer cells. Other target viral and bacterial mRNAs include, but are not limited to, those disclosed in Section 6.

5 One of skill in the art will appreciate that, although such target RNAs are functionally conserved in various species (*e.g.*, from yeast to humans), they exhibit nucleotide sequence and structural diversity. Therefore, inhibition of, for example, yeast telomerase by an anti-fungal compound identified by the methods of the invention might not interfere with human telomerase and normal human cell proliferation.

10 Thus, the methods of the invention can be used to identify test compounds that interfere with one or more target RNA interactions with host cell factors that are important for cell growth or viability, or essential in the life cycle of a virus, a bacterium, a protozoa or a fungus. Such test compounds and/or congeners that demonstrate desirable biologic and pharmacologic activity can be administered to a patient in need thereof in order to treat or prevent a disease caused by viral, bacterial, protozoan, or fungal infections. Such 15 diseases include, but are not limited to, HIV infection, AIDS, human T-cell leukemia, SIV infection, FIV infection, feline leukemia, hepatitis A, hepatitis B, hepatitis C, Dengue fever, malaria, rotavirus infection, severe acute gastroenteritis, diarrhea, encephalitis, hemorrhagic fever, syphilis, legionella, whooping cough, gonorrhea, sepsis, influenza, pneumonia, tinea infection, candida infection, and meningitis.

20 25 Non-limiting examples of RNA elements involved in the regulation of gene expression, *i.e.*, mRNA stability, translational efficiency via translational initiation and ribosome assembly, *etc.*, include the HIV TAR element, internal ribosome entry site, "slippery site", instability elements, and adenylate uridylate-rich elements, as discussed below.

5.1.1. HIV TAR Element

Transcriptional up-regulation of the genes of human immunodeficiency virus type 1 ("HIV-1") requires binding of the HIV Tat protein to the HIV trans-activation 30 response region RNA ("TAR RNA"), a 59-base stem-loop structure located at the 5' end of all nascent HIV-1 transcripts (Jones & Peterlin, 1994, *Annu. Rev. Biochem.* 63:717-43). Tat protein is known to interact with uracil 23 in the bulge region of the stem of TAR RNA. Thus, TAR RNA is a useful binding target for test compounds, such as small peptides and peptide analogs that bind to the bulge region of TAR RNA and inhibit formation of a Tat-TAR RNA complex involved in HIV-1 up-regulation (see Hwang *et al.*, 1999 *Proc. Natl. Acad. Sci. USA* 96:12997-13002). Accordingly, test compounds that bind to TAR RNA

can be useful as anti-HIV therapeutics (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Hamy *et al.*, 1998, Biochemistry 37:5086-5095; Mei *et al.*, 1998, Biochemistry 37:14204-14212), and therefore, are useful for treating or preventing AIDS.

5

5.1.2. Internal Ribosome Entry Site (“IRES”)

Internal ribosome entry sites (“IRES”) are found in the 5' untranslated regions (“5' UTR”) of several mRNAs, and are thought to be involved in the regulation of translational efficiency. When the IRES element is present on an mRNA downstream of a 10 translational stop codon, it directs ribosomal re-entry (Ghattas *et al.*, 1991, Mol. Cell. Biol. 11:5848-5959), which permits initiation of translation at the start of a second open reading frame.

As reviewed by Jang *et al.*, a large segment of the 5' nontranslated region, approximately 400 nucleotides in length, promotes internal entry of ribosomes independent 15 of the non-capped 5' end of picornavirus mRNAs (mammalian plus-strand RNA viruses whose genomes serve as mRNA). This 400 nucleotide segment (IRES), maps approximately 200 nt down-stream from the 5' end and is highly structured. IRES elements of different picornaviruses, although functionally similar *in vitro* and *in vivo*, are not identical in sequence or structure. However, IRES elements of the genera entero- and 20 rhinoviruses, on the one hand, and cardio- and aphthoviruses, on the other hand, reveal similarities corresponding to phylogenetic kinship. All IRES elements contain a conserved Yn-Xm-AUG unit (Y, pyrimidine; X, nucleotide) which appears essential for IRES function. The IRES elements of cardio-, entero- and aphthoviruses bind a cellular protein, p57. In the case of cardioviruses, the interaction between a specific stem-loop of the IREs is 25 essential for translation *in vitro*. The IRES elements of entero- and cardioviruses also bind the cellular protein, p52, but the significance of this interaction remains to be shown. The function of p57 or p52 in cellular metabolism is unknown. Since picornaviral IRES elements function *in vivo* in the absence of any viral gene products, it is speculated that IRES-like elements may also occur in specific cellular mRNAs releasing them from 30 cap-dependent translation (Jang *et al.*, 1990, Enzyme 44(1-4):292-309).

5.1.3. “Slippery Site”

Programmed, or directed, ribosomal frameshifting, when ribosomes shift from one translation reading frame to another and synthesize two viral proteins from a 35 single viral mRNA, is directed by a unique site in viral mRNAs called the “slippery site.” The slippery site directs ribosomal frameshifting in the -1 or +1 direction that causes the

ribosome to slip by one base in the 5' direction thereby placing the ribosome in the new reading frame to produce a new protein.

Programmed, or directed, ribosomal frameshifting is of particular value to viruses that package their plus strands, as it eliminates the need to splice their mRNAs and reduces the risk of packaging defective genomes and regulates the ratio of viral proteins synthesized. Examples of programmed translational frameshifting (both +1 and -1 shifts) have been identified in ScV systems (Lopinski *et al.*, 2000, Mol. Cell. Biol. 20(4):1095-103, retroviruses (Falk *et al.*, 1993, J. Virol. 67:273-6277; Jacks & Varmus, 1985, Science 230:1237-1242; Morikawa & Bishop, 1992, Virology 186:389-397; Nam *et al.*, 1993, J. Virol. 67:196-203); coronaviruses (Brierley *et al.*, 1987, EMBO J. 6:3779-3785; Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842); giardiviruses, which are also members of the Totiviridae (Wang *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:8595-8599); two bacterial genes (Blinkowa & Walker, 1990, Nucleic Acids Res., 18:1725-1729; Craigen & Caskey, 1986, Nature 322:273); bacteriophage genes (Condron *et al.*, 1991, Nucleic Acids Res. 19:5607-5612); astroviruses (Marczinke *et al.*, 1994, J. Virol. 68:5588-5595); the yeast EST3 gene (Lundblad & Morris, 1997, Curr. Biol. 7:969-976); and the rat, mouse, Xenopus, and Drosophila ornithine decarboxylase antizymes (Matsufuji *et al.*, 1995, Cell 80:51-60); and a significant number of cellular genes (Herold & Siddell, 1993, Nucleic Acids Res. 21:5838-5842).

Drugs targeted to ribosomal frameshifting minimize the problem of virus drug resistance because this strategy targets a host cellular process rather than one introduced into the cell by the virus, which minimizes the ability of viruses to evolve drug-resistant mutants. Compounds that target the RNA elements involved in regulating programmed frameshifting should have several advantages, including (a) any selective pressure on the host cellular translational machinery to adapt to the drugs would have to occur at the host evolutionary time scale, which is on the order of millions of years, (b) ribosomal frameshifting is not used to express any host proteins, and (c) altering viral frameshifting efficiencies by modulating the activity of a host protein minimizing the likelihood that the virus will acquire resistance to such inhibition by mutations in its own genome.

5.1.4. Instability Elements

"Instability elements" may be defined as specific sequence elements that promote the recognition of unstable mRNAs by cellular turnover machinery. Instability

elements have been found within mRNA protein coding regions as well as untranslated regions.

5 Altering the control of stability of normal mRNAs may lead to disease. The alteration of mRNA stability has been implicated in diseases such as, but not limited to, cancer, immune disorders, heart disease, and fibrotic disorders.

There are several examples of mutations that delete instability elements which then result in stabilization of mRNAs that may be involved in the onset of cancer. In Burkitt's lymphoma, a portion of the *c-myc* proto-oncogene is translocated to an Ig locus, producing a form of the *c-myc* mRNA that is five times more stable (see, e.g., Kapstein *et al.*, 1996, *J. Biol. Chem.* 271(31):18875-84). The highly oncogenic *v-fos* mRNA lacks the 10 3' UTR adenylate uridylate rich element ("ARE") that is found in the more labile and weakly oncogenic *c-fos* mRNA (see, e.g., Schiavi *et al.*, 1992, *Biochim Biophys Acta* 1114(2-3):95-106). Differences between the benign cervical lesions brought about by 15 nonintegrated circular human papillomavirus type 16 and its integrated form, that lacks the 3' UTR ARE and correlates with cervical carcinomas, may be a consequence of stabilizing the E6/E7 transcripts encoding oncogenic proteins. Integration of the virus results in deletion of the ARE instability element, resulting in stabilization of the transcripts and over-expression of the proteins (see, e.g., Jeon & Lambert, 1995, *Proc. Natl. Acad. Sci. USA* 92(5):1654-8). Deletion of AREs from the 3' UTR of the IL-2 and IL-3 genes promotes 20 increased stabilization of these mRNAs, high expression of these proteins, and leads to the formation of cancerous cells (see, e.g., Stoecklin *et al.*, 2000, *Mol. Cell. Biol.* 20(11):3753-63).

Mutations in trans-acting factors involved in mRNA turnover may also 25 promote cancer. In monocytic tumors, the lymphokine GM-CSF mRNA is specifically stabilized as a consequence of an oncogenic lesion in a trans-acting factor that controls mRNA turnover rates. Furthermore, the normally unstable IL-3 transcript is inappropriately long-lived in mast tumor cells. Similarly, the labile GM-CSF mRNA is greatly stabilized in bladder carcinoma cells. See, e.g., Bickel *et al.*, 1990, *J. Immunol.* 145(3):840-5.

30 The immune system is regulated by a large number of regulatory molecules that either activate or inhibit the immune response. It has now been clearly demonstrated that stability of the transcripts encoding these proteins are highly regulated. Altered regulation of these molecules leads to mis-regulation of this process and can result in drastic medical consequences. For example, recent results using transgenic mice have shown that 35 mis-regulation of the stability of the important modulator TNF α mRNA leads to diseases

such as, but not limited to, rheumatoid arthritis and a Crohn's-like liver disease. *See, e.g.*, Clark, 2000, *Arthritis Res.* 2(3):172-4.

Smooth muscle in the heart is modulated by the β -adrenergic receptor, which 5 in turn responds to the sympathetic neurotransmitter norepinephrine and the adrenal hormone epinephrine. Chronic heart failure is characterized by impairment of smooth muscle cells, which results, in part, from the more rapid decay of the β -adrenergic receptor mRNA. *See, e.g.*, Ellis & Frielle, 1999, *Biochem. Biophys. Res. Commun.* 258(3):552-8.

A large number of diseases result from over-expression of collagen. For 10 example, cirrhosis results from damage to the liver as a consequence of cancer, viral infection, or alcohol abuse. Such damage causes mis-regulation of collagen expression, leading to the formation of large collagen deposits. Recent results indicate that the sizeable increase in collagen expression is largely attributable to stabilization of its mRNA. *See, e.g.*, Lindquist *et al.*, 2000, *Am. J. Physiol. Gastrointest. Liver Physiol.* 279(3):G471-6.

15

5.1.5. Adenylate Uridylate-rich Elements (“ARE”)

Adenylate uridylate-rich elements (“ARE”) are found in the 3' untranslated regions (“3' UTR”) of several mRNAs, and involved in the turnover of mRNAs, such as but not limited to transcription factors, cytokines, and lymphokines. AREs may function both 20 as stabilizing and destabilizing elements. ARE mRNAs are classified into five groups, depending on sequence (Bakheet *et al.*, 2001, *Nucl. Acids Res.* 29(1):246-254). An ongoing database at the web site <http://rc.kfshrc.edu.sa/ared> contains ARE-containing mRNAs and their cluster groups, which is incorporated by reference in its entirety. The ARE motifs are classified as follows:

25	Group I Cluster	(AUUUAUUUAUUUAUUUAUUU)	SEQ ID NO: 1
	Group II Cluster	(AUUUAUUUAUUUAUUU) stretch	SEQ ID NO: 2
	Group III Cluster	(WAUUUAUUUAUUUAW) stretch	SEQ ID NO: 3
	Group IV Cluster	(WWAUUUAUUUAWW) stretch	SEQ ID NO: 4
30	Group V Cluster	(WWWWAUUUAWWWW) stretch	SEQ ID NO: 5

The ARE-mRNAs were clustered into five groups containing five, four, three and two pentameric repeats, while the last group contains only one pentamer within the 13-bp ARE pattern. Functional categories were assigned whenever possible according to NCBI-COG functional annotation (Tatusov *et al.*, 2001, *Nucleic Acids Research*, 29(1): 35 22-28), in addition to the categories: inflammation, immune response, development/differentiation, using an extensive literature search.

Group I contains many secreted proteins including GM-CSF, IL-1, IL-11, IL-12 and Gro- β that affect the growth of hematopoietic and immune cells (Witsell & Schook, 1992, Proc. Natl Acad. Sci. USA, 89:4754–4758). Although TNF α is both a pro-inflammatory and anti-tumor protein, there is experimental evidence that it can act as a growth factor in certain leukemias and lymphomas (Liu *et al.*, 2000, J. Biol. Chem. 275:21086–21093).

Unlike Group I, Groups II–V contain functionally diverse gene families comprising immune response, cell cycle and proliferation, inflammation and coagulation, angiogenesis, metabolism, energy, DNA binding and transcription, nutrient transportation and ionic homeostasis, protein synthesis, cellular biogenesis, signal transduction, and apoptosis (Bakheet *et al.*, 2001, Nucl. Acids Res. 29(1):246-254).

Several groups have described ARE-binding proteins that influence the ARE-mRNA stability. Among the well-characterized proteins are the mammalian homologs of ELAV (embryonic lethal abnormal vision) proteins including AUF1, HuR and He1-N2 (Zhang *et al.*, 1993, Mol. Cell. Biol. 13:7652–7665; Levine *et al.*, 1993, Mol. Cell. Biol. 13:3494–3504; Ma *et al.*, 1996, J. Biol. Chem. 271:8144–8151). The zinc-finger protein tristetraprolin has been identified as another ARE-binding protein with destabilizing activity on TNF α , IL-3 and GM-CSF mRNAs (Stoecklin *et al.*, 2000, Mol. Cell. Biol. 20:3753–3763; Carballo *et al.*, 2000, Blood 95:1891–1899).

Since ARE-containing genes are clearly important in biological systems, including but not limited to a number of the early response genes that regulate cell proliferation and responses to exogenous agents, the identification of compounds that bind to one or more of the ARE clusters and potentially modulate the stability of the target RNA can potentially be of value as a therapeutic.

5.2. Detectably Labeled Target RNAs

Target nucleic acids, including but not limited to RNA and DNA, useful in the methods of the present invention have a label that is detectable via conventional spectroscopic means or radiographic means. Preferably, target nucleic acids are labeled with a covalently attached dye molecule. Useful dye-molecule labels include, but are not limited to, fluorescent dyes, phosphorescent dyes, ultraviolet dyes, infrared dyes, and visible dyes. Preferably, the dye is a visible dye.

Useful labels in the present invention can include, but are not limited to, spectroscopic labels such as fluorescent dyes (*e.g.*, fluorescein and derivatives such as fluorescein isothiocyanate (FITC) and Oregon Green™, rhodamine and derivatives (*e.g.*,

Texas red, tetramethylrhodamine isothiocyanate (TRITC), bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, *etc.*), digoxigenin, biotin, phycoerythrin, AMCA, CyDye™, and the like), radiolabels (*e.g.*, ^3H , ^{125}I , ^{35}S , ^{14}C , ^{32}P , ^{33}P , *etc.*), enzymes (*e.g.*, horse radish peroxidase, alkaline phosphatase *etc.*), spectroscopic colorimetric labels such as colloidal gold or colored glass or plastic (*e.g.* polystyrene, polypropylene, latex, *etc.*) beads, or nanoparticles – nanoclusters of inorganic ions with defined dimension from 0.1 to 1000 nm. Useful affinity tags and complimentary partners include, but are not limited to, biotin-streptavidin, complimentary nucleic acid fragments (*e.g.*, oligo dT-oligo dA, oligo T-oligo A, oligo dG-oligo dC, oligo G-oligo C), aptamer-streptavidin, or haptens and proteins for which antisera or monoclonal antibodies are available. The label may be coupled directly or indirectly to a component of the detection assay (*e.g.*, the detection reagent) according to methods well known in the art. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions.

In one embodiment, nucleic acids that are labeled at one or more specific locations are chemically synthesized using phosphoramidite or other solution or solid-phase methods. Detailed descriptions of the chemistry used to form polynucleotides by the phosphoramidite method are well known (*see, e.g.*, Caruthers *et al.*, U.S. Pat. Nos. 4,458,066 and 4,415,732; Caruthers *et al.*, 1982, Genetic Engineering 4:1-17; *Users Manual Model 392 and 394 Polynucleotide Synthesizers*, 1990, pages 6-1 through 6-22, Applied Biosystems, Part No. 901237; Ojwang, *et al.*, 1997, Biochemistry, 36:6033-6045). The phosphoramidite method of polynucleotide synthesis is the preferred method because of its efficient and rapid coupling and the stability of the starting materials. The synthesis is performed with the growing polynucleotide chain attached to a solid support, such that excess reagents, which are generally in the liquid phase, can be easily removed by washing, decanting, and/or filtration, thereby eliminating the need for purification steps between synthesis cycles.

The following briefly describes illustrative steps of a typical polynucleotide synthesis cycle using the phosphoramidite method. First, a solid support to which is attached a protected nucleoside monomer at its 3' terminus is treated with acid, *e.g.*, trichloroacetic acid, to remove the 5'-hydroxyl protecting group, freeing the hydroxyl group for a subsequent coupling reaction. After the coupling reaction is completed an activated intermediate is formed by contacting the support-bound nucleoside with a protected nucleoside phosphoramidite monomer and a weak acid, *e.g.*, tetrazole. The weak acid protonates the nitrogen atom of the phosphoramidite forming a reactive intermediate.

Nucleoside addition is generally complete within 30 seconds. Next, a capping step is performed, which terminates any polynucleotide chains that did not undergo nucleoside addition. Capping is preferably performed using acetic anhydride and 1-methylimidazole.

5 The phosphite group of the internucleotide linkage is then converted to the more stable phosphotriester by oxidation using iodine as the preferred oxidizing agent and water as the oxygen donor. After oxidation, the hydroxyl protecting group of the newly added nucleoside is removed with a protic acid, e.g., trichloroacetic acid or dichloroacetic acid, and the cycle is repeated one or more times until chain elongation is complete. After

10 synthesis, the polynucleotide chain is cleaved from the support using a base, e.g., ammonium hydroxide or t-butyl amine. The cleavage reaction also removes any phosphate protecting groups, e.g., cyanoethyl. Finally, the protecting groups on the exocyclic amines of the bases and any protecting groups on the dyes are removed by treating the polynucleotide solution in base at an elevated temperature, e.g., at about 55°C. Preferably

15 the various protecting groups are removed using ammonium hydroxide or t-butyl amine.

Any of the nucleoside phosphoramidite monomers can be labeled using standard phosphoramidite chemistry methods (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002; Ojwang *et al.*, 1997, Biochemistry. 36:6033-6045 and references cited therein). Dye molecules useful for covalently coupling to phosphoramidites preferably comprise a primary hydroxyl group that is not part of the dye's chromophore. Illustrative dye molecules include, but are not limited to, disperse dye CAS 4439-31-0, disperse dye CAS 6054-58-6, disperse dye CAS 4392-69-2 (Sigma-Aldrich, St. Louis, MO), disperse red, and 1-pyrenebutanol (Molecular Probes, Eugene, OR). Other dyes useful for coupling to phosphoramidites will be apparent to those of skill in the art, such as fluorescein, cy3, and cy5 fluorescent dyes, and may be purchased from, e.g., Sigma-Aldrich, St. Louis, MO or Molecular Probes, Inc., Eugene, OR.

In another embodiment, dye-labeled target RNA molecules are synthesized enzymatically using *in vitro* transcription (Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). In this embodiment, a template DNA is denatured by heating to about 90°C and an oligonucleotide primer is annealed to the template DNA, for example by slow-cooling the mixture of the denatured template and the primer from about 90°C to room temperature. A mixture of ribonucleoside-5'-triphosphates capable of supporting template-directed enzymatic extension of the primed template (e.g., a mixture including GTP, ATP, CTP, and UTP), including one or more dye-labeled ribonucleotides (Sigma-Aldrich, St. Louis, MO), is added to the primed template. Next, a polymerase enzyme is added to the mixture under conditions where the polymerase enzyme

is active, which are well-known to those skilled in the art. A labeled polynucleotide is formed by the incorporation of the labeled ribonucleotides during polymerase-mediated strand synthesis.

In yet another embodiment of the invention, nucleic acid molecules are end-labeled after their synthesis. Methods for labeling the 5'-end of an oligonucleotide include but are by no means limited to: (i) periodate oxidation of a 5'-to-5'-coupled ribonucleotide, followed by reaction with an amine-reactive label (Heller & Morisson, 1985, in *Rapid Detection and Identification of Infectious Agents*, D.T. Kingsbury and S. Falkow, eds., pp. 245-256, Academic Press); (ii) condensation of ethylenediamine with 5'-phosphorylated polynucleotide, followed by reaction with an amine reactive label (Morrison, European Patent Application 232 967); (iii) introduction of an aliphatic amine substituent using an aminohexyl phosphite reagent in solid-phase DNA synthesis, followed by reaction with an amine reactive label (Cardullo *et al.*, 1988, Proc. Natl. Acad. Sci. USA 85:8790-8794); and (iv) introduction of a thiophosphate group on the 5'-end of the nucleic acid, using phosphatase treatment followed by end-labeling with ATP-³²S and kinase, which reacts specifically and efficiently with maleimide-labeled fluorescent dyes (Czworkowski *et al.*, 1991, Biochem. 30:4821-4830).

A detectable label should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site. Accordingly, if the region of the target nucleic acid that binds to a host cell factor is known, a detectable label is preferably incorporated into the nucleic acid molecule at one or more positions that are spatially or sequentially remote from the binding region.

After synthesis, the labeled target nucleic acid can be purified using standard techniques known to those skilled in the art (see Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96(23):12997-13002 and references cited therein). Depending on the length of the target nucleic acid and the method of its synthesis, such purification techniques include, but are not limited to, reverse-phase high-performance liquid chromatography ("reverse-phase HPLC"), fast performance liquid chromatography ("FPLC"), and gel purification. After purification, the target RNA is refolded into its native conformation, preferably by heating to approximately 85-95°C and slowly cooling to room temperature in a buffer, e.g., a buffer comprising about 50 mM Tris-HCl, pH 8 and 100 mM NaCl.

In another embodiment, the target nucleic acid can also be radiolabeled. A radiolabel, such as, but not limited to, an isotope of phosphorus, sulfur, or hydrogen, may be

incorporated into a nucleotide, which is added either after or during the synthesis of the target nucleic acid. Methods for the synthesis and purification of radiolabeled nucleic acids are well known to one of skill in the art. See, e.g., Sambrook *et al.*, 1989, in *Molecular Cloning: A Laboratory Manual*, pp 10.2-10.70, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties.

In another embodiment, the target nucleic acid can be attached to an inorganic nanoparticle. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag₂S, ZnS, CdS, CdTe, Au, or TiO₂. Nanoparticles have unique optical, electronic and catalytic properties relative to bulk materials which can be adjusted according to the size of the particle. Methods for the attachment of nucleic acids are well known to one of skill in the art (see, e.g., Niemeyer, 2001, Angew. Chem. Int. Ed. 40: 4129-4158, International Patent Publication WO/0218643, and the references cited therein, the disclosures of which are hereby incorporated by reference in their entireties).

5.3. Libraries of Small Molecules

Libraries screened using the methods of the present invention can comprise a variety of types of test compounds. In some embodiments, the test compounds are nucleic acid or peptide molecules. In a non-limiting example, peptide molecules can exist in a phage display library. In other embodiments, types of test compounds include, but are not limited to, peptide analogs including peptides comprising non-naturally occurring amino acids, e.g., D-amino acids, phosphorous analogs of amino acids, such as α-amino phosphoric acids and α-amino phosphoric acids, or amino acids having non-peptide linkages, nucleic acid analogs such as phosphorothioates and PNAs, hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose. Libraries of polypeptides or proteins can also be used.

In a preferred embodiment, the combinatorial libraries are small organic molecule libraries, such as, but not limited to, benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, and diazepindiones. In another embodiment, the combinatorial libraries comprise peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinyllogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; or carbohydrate libraries.

Combinatorial libraries are themselves commercially available (see, e.g., Advanced ChemTech Europe Ltd., Cambridgeshire, UK; ASINEX, Moscow Russia; BioFocus plc, Sittingbourne, UK; Bionet Research (A division of Key Organics Limited), Camelford, UK; ChemBridge Corporation, San Diego, California; ChemDiv Inc, San Diego, California.; ChemRx Advanced Technologies, South San Francisco, California; ComGenex Inc., Budapest, Hungary; Evotec OAI Ltd, Abingdon, UK; IF LAB Ltd., Kiev, Ukraine; Maybridge plc, Cornwall, UK; PharmaCore, Inc., North Carolina; SIDDCO Inc, Tucson, Arizona; TimTec Inc, Newark, Delaware; Tripos Receptor Research Ltd, Bude, UK; Toslab, Ekaterinburg, Russia).

In one embodiment, the combinatorial compound library for the methods of the present invention may be synthesized. There is a great interest in synthetic methods directed toward the creation of large collections of small organic compounds, or libraries, which could be screened for pharmacological, biological or other activity (Dolle, 2001, J. Comb. Chem. 3:477-517; Hall *et al.*, 2001, J. Comb. Chem. 3:125-150; Dolle, 2000, J. Comb. Chem. 2:383-433; Dolle, 1999, J. Comb. Chem. 1:235-282). The synthetic methods applied to create vast combinatorial libraries are performed in solution or in the solid phase, *i.e.*, on a solid support. Solid-phase synthesis makes it easier to conduct multi-step reactions and to drive reactions to completion with high yields because excess reagents can be easily added and washed away after each reaction step. Solid-phase combinatorial synthesis also tends to improve isolation, purification and screening. However, the more traditional solution phase chemistry supports a wider variety of organic reactions than solid-phase chemistry. Methods and strategies for the synthesis of combinatorial libraries can be found in *A Practical Guide to Combinatorial Chemistry*, A.W. Czarnik and S.H. Dewitt , eds., American Chemical Society, 1997; *The Combinatorial Index*, B.A. Bunin, Academic Press, 1998; *Organic Synthesis on Solid Phase*, F.Z. Dörwald, Wiley-VCH, 2000; and *Solid-Phase Organic Syntheses*, Vol. 1, A.W. Czarnik, ed., Wiley Interscience, 2001.

Combinatorial compound libraries of the present invention may be synthesized using apparatuses described in US Patent No. 6,358,479 to Frisina *et al.*, U.S. Patent No. 6,190,619 to Kilcoin *et al.*, US Patent No. 6,132,686 to Gallup *et al.*, US Patent No. 6,126,904 to Zuellig *et al.*, US Patent No. 6,074,613 to Harness *et al.*, US Patent No. 6,054,100 to Stanchfield *et al.*, and US Patent No. 5,746,982 to Saneii *et al.* which are hereby incorporated by reference in their entirety. These patents describe synthesis apparatuses capable of holding a plurality of reaction vessels for parallel synthesis of multiple discrete compounds or for combinatorial libraries of compounds.

In one embodiment, the combinatorial compound library can be synthesized in solution. The method disclosed in U.S. Patent No. 6,194,612 to Boger *et al.*, which is hereby incorporated by reference in its entirety, features compounds useful as templates for solution phase synthesis of combinatorial libraries. The template is designed to permit reaction products to be easily purified from unreacted reactants using liquid/liquid or solid/liquid extractions. The compounds produced by combinatorial synthesis using the template will preferably be small organic molecules. Some compounds in the library may mimic the effects of non-peptides or peptides. In contrast to solid phase synthesize of combinatorial compound libraries, liquid phase synthesis does not require the use of specialized protocols for monitoring the individual steps of a multistep solid phase synthesis (Egner *et al.*, 1995, J.Org. Chem. 60:2652; Anderson *et al.*, 1995, J. Org. Chem. 60:2650; Fitch *et al.*, 1994, J. Org. Chem. 59:7955; Look *et al.*, 1994, J. Org. Chem. 49:7588; Metzger *et al.*, 1993, Angew. Chem., Int. Ed. Engl. 32:894; Youngquist *et al.*, 1994, Rapid Commun. Mass Spect. 8:77; Chu *et al.*, 1995, J. Am. Chem. Soc. 117:5419; Brummel *et al.*, 1994, Science 264:399; Stevanovic *et al.*, 1993, Bioorg. Med. Chem. Lett. 3:431).

Combinatorial compound libraries useful for the methods of the present invention can be synthesized on solid supports. In one embodiment, a split synthesis method, a protocol of separating and mixing solid supports during the synthesis, is used to synthesize a library of compounds on solid supports (see Lam *et al.*, 1997, Chem. Rev. 97:41-448; Ohlmeyer *et al.*, 1993, Proc. Natl. Acad. Sci. USA 90:10922-10926 and references cited therein). Each solid support in the final library has substantially one type of test compound attached to its surface. Other methods for synthesizing combinatorial libraries on solid supports, wherein one product is attached to each support, will be known to those of skill in the art (see, e.g., Nefzi *et al.*, 1997, Chem. Rev. 97:449-472 and US Patent No. 6,087,186 to Cargill *et al.* which are hereby incorporated by reference in their entirety).

As used herein, the term "solid support" is not limited to a specific type of solid support. Rather a large number of supports are available and are known to one skilled in the art. Solid supports include silica gels, resins, derivatized plastic films, glass beads, cotton, plastic beads, polystyrene beads, alumina gels, and polysaccharides. A suitable solid support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, a solid support can be a resin such as p-methylbenzhydrylamine (pMBHA) resin (Peptides International, Louisville, KY), polystyrenes (e.g., PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), including chloromethylpolystyrene, hydroxymethylpolystyrene and

aminomethylpolystyrene, poly (dimethylacrylamide)-grafted styrene co-divinyl-benzene (e.g., POLYHIPE resin, obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (e.g., TENTAGEL or ARGOGEL, Bayer, Tubingen, Germany) polydimethylacrylamide resin (obtained from Milligen/Bioscience, California), or Sepharose (Pharmacia, Sweden).

In one embodiment, the solid phase support is suitable for *in vivo* use, i.e., it can serve as a carrier or support for administration of the test compound to a patient (e.g., TENTAGEL, Bayer, Tubingen, Germany). In a particular embodiment, the solid support is palatable and/or orally ingestable.

In some embodiments of the present invention, compounds can be attached to solid supports via linkers. Linkers can be integral and part of the solid support, or they may be nonintegral that are either synthesized on the solid support or attached thereto after synthesis. Linkers are useful not only for providing points of test compound attachment to the solid support, but also for allowing different groups of molecules to be cleaved from the solid support under different conditions, depending on the nature of the linker. For example, linkers can be, *inter alia*, electrophilically cleaved, nucleophilically cleaved, photocleavable, enzymatically cleaved, cleaved by metals, cleaved under reductive conditions or cleaved under oxidative conditions.

In another embodiment, the combinatorial compound libraries can be assembled *in situ* using dynamic combinatorial chemistry as described in European Patent Application 1,118,359 A1 to Lehn; Huc & Nguyen, 2001, Comb. Chem. High Throughput. Screen. 4:53-74; Lehn and Eliseev, 2001, Science 291:2331-2332; Cousins *et al.* 2000, Curr. Opin. Chem. Biol. 4: 270-279; and Karan & Miller, 2000, Drug. Disc. Today 5:67-75 which are incorporated by reference in their entirety.

Dynamic combinatorial chemistry uses non-covalent interaction with a target biomolecule, including but not limited to a protein, RNA, or DNA, to favor assembly of the most tightly binding molecule that is a combination of constituent subunits present as a mixture in the presence of the biomolecule. According to the laws of thermodynamics, when a collection of molecules is able to combine and recombine at equilibrium through reversible chemical reactions in solution, molecules, preferably one molecule, that bind most tightly to a templating biomolecule will be present in greater amount than all other possible combinations. The reversible chemical reactions include, but are not limited to, imine, acyl-hydrazone, amide, acetal, or ester formation between carbonyl-containing compounds and amines, hydrazines, or alcohols; thiol exchange between disulfides; alcohol

exchange in borate esters; Diels-Alder reactions; thermal- or photoinduced sigmatropic or electrocyclic rearrangements; or Michael reactions.

In the preferred embodiment of this technique, the constituent components of the dynamic combinatorial compound library are allowed to combine and reach equilibrium in the absence of the target RNA and then incubated in the presence of the target RNA, preferably at physiological conditions, until a second equilibrium is reached. The second, perturbed, equilibrium (the so-called "templated mixture") can, but need not necessarily, be fixed by a further chemical transformation, including but not limited to reduction, oxidation, hydrolysis, acidification, or basification, to prevent restoration of the original equilibrium when the dynamical combinatorial compound library is separated from the target RNA.

In the preferred embodiment of this technique, the predominant product or products of the templated dynamic combinatorial library can be separated from the minor products and directly identified. In another embodiment, the identity of the predominant product or products can be identified by a deconvolution strategy involving preparation of derivative dynamic combinatorial libraries, as described in European Patent Application 1,118,359 A1, which is incorporated by reference in their entirety, whereby each component of the mixture is, preferably one-by-one but possibly group-wise, left out of the mixture and the ability of the derivative library mixture at chemical equilibrium to bind the target RNA is measured. The components whose removal most greatly reduces the ability of the derivative dynamic combinatorial library to bind the target RNA are likely the components of the predominant product or products in the original dynamic combinatorial library.

25

5.4. Library Screening

After a target nucleic acid, such as but not limited to RNA or DNA, is labeled and a test compound library is synthesized or purchased or both, the labeled target nucleic acid is used to screen the library to identify test compounds that bind to the nucleic acid. Screening comprises contacting a labeled target nucleic acid with an individual, or small group, of the components of the compound library. Preferably, the contacting occurs in an aqueous solution, and most preferably, under physiologic conditions. The aqueous solution preferably stabilizes the labeled target nucleic acid and prevents denaturation or degradation of the nucleic acid without interfering with binding of the test compounds. The aqueous solution can be similar to the solution in which a complex between the target RNA and its corresponding host cell factor (if known) is formed *in vitro*. For example, TK.

buffer, which is commonly used to form Tat protein-TAR RNA complexes *in vitro*, can be used in the methods of the invention as an aqueous solution to screen a library of test compounds for TAR RNA binding compounds.

5 The methods of the present invention for screening a library of test compounds preferably comprise contacting a test compound with a target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions. The aqueous solution optionally further comprises non-specific nucleic acids, such as, but not limited to, DNA; yeast tRNA; salmon sperm DNA; homoribopolymers such as, but not 10 limited to, poly IC, polyA, polyU, and polyC; and non-specific RNA. The non-specific RNA may be an unlabeled target nucleic acid having a mutation at the binding site, which renders the unlabeled nucleic acid incapable of interacting with a test compound at that site. For example, if dye-labeled TAR RNA is used to screen a library, unlabeled TAR RNA 15 having a mutation in the uracil 23/cytosine 24 bulge region may also be present in the aqueous solution. Without being bound by any theory, the addition of unlabeled RNA that is essentially identical to the dye-labeled target RNA except for a mutation at the binding site might minimize interactions of other regions of the dye-labeled target RNA with test compounds or with the solid support and prevent false positive results.

20 The solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. The pH of the solution typically ranges from about 5 to about 8, preferably from about 6 to about 8, most preferably from about 6.5 to about 8. A variety of buffers may be used to achieve the desired pH. Suitable buffers include, but are not limited to, Tris, Mes, Bis-Tris, Ada, Aces, Pipes, Mopso, Bis-Tris propane, Bes, 25 Mops, Tes, Hepes, Dipso, Mobs, Tapso, Trizma, Heppso, Popso, TEA, Epps, Tricine, Gly-Gly, Bicine, and sodium-potassium phosphate. The buffering agent comprises from about 10 mM to about 100 mM, preferably from about 25 mM to about 75 mM, most preferably from about 40 mM to about 60 mM buffering agent. The pH of the aqueous solution can be optimized for different screening reactions, depending on the target RNA used and the 30 types of test compounds in the library, and therefore, the type and amount of the buffer used in the solution can vary from screen to screen. In a preferred embodiment, the aqueous solution has a pH of about 7.4, which can be achieved using about 50 mM Tris buffer.

35 In addition to an appropriate buffer, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. Without

being bound by any theory, Applicant has found that a combination of KCl, NaCl, and MgCl₂ stabilizes the target RNA such that most of the RNA is not denatured or digested over the course of the screening reaction. The optional concentration of each salt used in the aqueous solution is dependent on the particular target RNA used and can be determined using routine experimentation.

The solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant. Without being bound by any theory, a small amount of detergent or surfactant in the solution might reduce non-specific binding of the target RNA to the solid support and control aggregation and increase stability of target RNA molecules.

Typical detergents useful in the methods of the present invention include, but are not limited to, anionic detergents, such as salts of deoxycholic acid, 1-heptanesulfonic acid, N-laurylsarcosine, lauryl sulfate, 1-octane sulfonic acid and taurocholic acid; cationic detergents such as benzalkonium chloride, cetylpyridinium, methylbenzethonium chloride, and decamethonium bromide; zwitterionic detergents such as CHAPS, CHAPSO, alkyl betaines, alkyl amidoalkyl betaines, N-dodecyl-N,N-dimethyl-3-ammonio-1-propanesulfonate, and phosphatidylcholine; and non-ionic detergents such as n-decyl α-D-glucopyranoside, n-decyl β-D-maltopyranoside, n-dodecyl β-D-maltoside, n-octyl β-D-glucopyranoside, sorbitan esters, n-tetradecyl β-D-maltoside, octylphenoxy polyethoxyethanol (Nonidet P-40), nonylphenoxypolyethoxyethanol (NP-40), and tritons.

Preferably, the detergent, if present, is a nonionic detergent. Typical surfactants useful in the methods of the present invention include, but are not limited to, ammonium lauryl sulfate, polyethylene glycols, butyl glucoside, decyl glucoside, Polysorbate 80, lauric acid, myristic acid, palmitic acid, potassium palmitate, undecanoic acid, lauryl betaine, and lauryl alcohol. More preferably, the detergent, if present, is Triton X-100 and present in an amount of about 0.1% (w/v).

Non-specific binding of a labeled target nucleic acid to test compounds can be further minimized by treating the binding reaction with one or more blocking agents. In one embodiment, the binding reactions are treated with a blocking agent, e.g., bovine serum albumin ("BSA"), before contacting with the labeled target nucleic acid. In another embodiment, the binding reactions are treated sequentially with at least two different blocking agents. This blocking step is preferably performed at room temperature for from about 0.5 to about 3 hours. In a subsequent step, the reaction mixture is further treated with unlabeled RNA having a mutation at the binding site. This blocking step is preferably performed at about 4°C for from about 12 hours to about 36 hours before addition of the dye-labeled target RNA. Preferably, the solution used in the one or more blocking steps is

substantially similar to the aqueous solution used to screen the library with the dye-labeled target RNA, *e.g.*, in pH and salt concentration.

Once contacted, the mixture of labeled target nucleic acid and the test compound is preferably maintained at 4°C for from about 1 day to about 5 days, preferably from about 2 days to about 3 days with constant agitation. To identify the reactions in which binding to the labeled target nucleic acid occurred, after the incubation period, bound from free compounds are determined using an electrophoretic technique (see Section 5.5.1), or any of the methods disclosed in Section 5.5 *infra*. In another embodiment, the complexed target nucleic acid does not need to be separated from the free target nucleic acid if a technique (*i.e.*, spectrometry) that differentiates between bound and unbound target nucleic acids is used.

The methods for identifying small molecules bound to labeled nucleic acid will vary with the type of label on the target nucleic acid. For example, if a target RNA is labeled with a visible or fluorescent dye, the target RNA complexes are preferably identified using a chromatographic technique that separates bound from free target by an electrophoretic or size differential technique using individual reactions. The reactions corresponding to changes in the migration of the complexed RNA can be cross-referenced to the small molecule compound(s) added to said reaction. Alternatively, complexed target RNA can be screened *en masse* and then separated from free target RNA using an electrophoretic or size differential technique, the resultant complexed target is then analyzed using a mass spectrometric technique. In this fashion the bound small molecule can be identified on the basis of its molecular weight. In this reaction *a priori* knowledge of the exact molecular weights of all compounds within the library is known. In another embodiment, the test compounds bound to the target nucleic acid may not require separation from the unbound target nucleic acid if a technique such as, but not limited to, spectrometry is used.

5.5. Separation Methods for Screening Test Compounds

Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids. Methods that can be utilized for the physical separation of complexed target RNA from unbound target RNA include, but are not limited to, electrophoresis, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography,

and nanoparticle aggregation.

5.5.1. Electrophoresis

Methods for separation of the complex of a target RNA bound to a test
5 compound from the unbound RNA comprises any method of electrophoretic separation,
including but not limited to, denaturing and non-denaturing polyacrylamide gel
electrophoresis, urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two
dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis,
agarose gel electrophoresis, and capillary electrophoresis.
10

In a preferred embodiment, an automated electrophoretic system comprising
a capillary cartridge having a plurality of capillary tubes is used for high-throughput
screening of test compounds bound to target RNA. Such an apparatus for performing
automated capillary gel electrophoresis is disclosed in U.S. Patent Nos. 5,885,430;
15 5,916,428; 6,027,627; and 6,063,251, the disclosures of which are incorporated by
reference in their entireties.

The device disclosed in U.S. Patent No. 5,885,430, which is incorporated by
reference in its entirety, allows one to simultaneously introduce samples into a plurality of
capillary tubes directly from microtiter trays having a standard size. U.S. Patent No.
20 5,885,430 discloses a disposable capillary cartridge which can be cleaned between
electrophoresis runs, the cartridge having a plurality of capillary tubes. A first end of each
capillary tube is retained in a mounting plate, the first ends collectively forming an array in
the mounting plate. The spacing between the first ends corresponds to the spacing between
the centers of the wells of a microtiter tray having a standard size. Thus, the first ends of
25 the capillary tubes can simultaneously be dipped into the samples present in the tray's wells.
The cartridge is provided with a second mounting plate in which the second ends of the
capillary tubes are retained. The second ends of the capillary tubes are arranged in an array
which corresponds to the wells in the microtiter tray, which allows for each capillary tube
to be isolated from its neighbors and therefore free from cross-contamination, as each end
is dipped into an individual well.
30

Plate holes may be provided in each mounting plate and the capillary tubes
inserted through these plate holes. In such a case, the plate holes are sealed airtight so that
the side of the mounting plate having the exposed capillary ends can be pressurized.
Application of a positive pressure in the vicinity of the capillary openings in this mounting
35 plate allows for the introduction of air and fluids during electrophoretic operations and also
can be used to force out gel and other materials from the capillary tubes during

reconditioning. The capillary tubes may be protected from damage using a needle comprising a cannula and/or plastic tubes, and the like when they are placed in these plate holes. When metallic cannula or the like are used, they can serve as electrical contacts for current flow during electrophoresis. In the presence of a second mounting plate, the second mounting plate is provided with plate holes through which the second ends of the capillary tubes project. In this instance, the second mounting plate serves as a pressure containment member of a pressure cell and the second ends of the capillary tubes communicate with an internal cavity of the pressure cell. The pressure cell is also formed with an inlet and an outlet. Gels, buffer solutions, cleaning agents, and the like may be introduced into the internal cavity through the inlet, and each of these can simultaneously enter the second ends of the capillaries.

In another preferred embodiment, the automated electrophoretic system can comprise a chip system consisting of complex designs of interconnected channels that perform and analyze enzyme reactions using part of a channel design as a tiny, continuously operating electrophoresis material, where reactions with one sample are going on in one area of the chip while electrophoretic separation of the products of another sample is taking place in a different part of the chip. Such a system is disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 5,942,443; 5,948,227; 6,042,709; 6,042,710; 15 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 20 6,171,850; and 6,186,660, the disclosures of which are incorporated by reference in their entireties.

The system disclosed in U.S. Patent No. 5,699,157, which is hereby incorporated by reference in its entirety, provides for a microfluidic system for high-speed electrophoretic analysis of subject materials for applications in the fields of chemistry, 25 biochemistry, biotechnology, molecular biology and numerous other areas. The system has a channel in a substrate, a light source and a photoreceptor. The channel holds subject materials in solution in an electric field so that the materials move through the channel and separate into bands according to species. The light source excites fluorescent light in the species bands and the photoreceptor is arranged to receive the fluorescent light from the bands. The system further has a means for masking the channel so that the photoreceptor can receive the fluorescent light only at periodically spaced regions along the channel. The system also has an unit connected to analyze the modulation frequencies of light intensity received by the photoreceptor so that velocities of the bands along the channel are 30 determined, which allows the materials to be analyzed.

35 The system disclosed in U.S. Patent No. 5,699,157 also provides for a

method of performing high-speed electrophoretic analysis of subject materials, which comprises the steps of holding the subject materials in solution in a channel of a microfluidic system; subjecting the materials to an electric field so that the subject materials move through the channel and separate into species bands; directing light toward the channel; receiving light from periodically spaced regions along the channel simultaneously; and analyzing the frequencies of light intensity of the received light so that velocities of the bands along the channel can be determined for analysis of said materials. The determination of the velocity of a species band determines the electrophoretic mobility of the species and its identification.

U.S. Patent No. 5,842,787, which is hereby incorporated by reference in its entirety, is generally directed to devices and systems employ channels having, at least in part, depths that are varied over those which have been previously described (such as the device disclosed in U.S. Patent No. 5,699,157), wherein said channel depths provide numerous beneficial and unexpected results such as but not limited to, a reduction in sample perturbation, reduced non-specific sample mixture by diffusion, and increased resolution.

In another embodiment, the electrophoretic method of separation comprises polyacrylamide gel electrophoresis. In a preferred embodiment, the polyacrylamide gel electrophoresis is non-denaturing, so as to differentiate the mobilities of the target RNA bound to a test compound from free target RNA. If the polyacrylamide gel electrophoresis is denaturing, then the target RNA:test compound complex must be cross-linked prior to electrophoresis to prevent the disassociation of the target RNA from the test compound during electrophoresis. Such techniques are well known to one of skill in the art.

In one embodiment of the method, the binding of test compounds to target nucleic acid can be detected, preferably in an automated fashion, by gel electrophoretic analysis of interference footprinting. RNA can be degraded at specific base sites by enzymatic methods such as ribonucleases A, U₂, CL₃, T₁, Phy M, and *B. cereus* or chemical methods such as diethylpyrocarbonate, sodium hydroxide, hydrazine, piperidine formate, dimethyl sulfate,

[2,12-dimethyl-3,7,11,17-tetraazacyclo[11.3.1]heptadeca-1(17),2,11,13,15-pentaenato] nickel(II) (NiCR), cobalt(II)chloride, or iron(II) ethylenediaminetetraacetate (Fe-EDTA) as described for example in Zheng *et al.*, 1999, Biochem. 37:2207-2214; Latham & Cech, 1989, Science 245:276-282; and Sambrook *et al.*, 2001, in Molecular Cloning: A Laboratory Manual, pp 12.61-12.73, Cold Spring Harbor Laboratory Press, and the references cited therein, which are hereby incorporated by reference in their entireties. The

specific pattern of cleavage sites is determined by the accessibility of particular bases to the reagent employed to initiate cleavage and, as such, is therefore determined by the three-dimensional structure of the RNA.

5 The interaction of small molecules with a target nucleic acid can change the accessibility of bases to these cleavage reagents both by causing conformational changes in the target nucleic acid or by covering a base at the binding interface. When a test compound binds to the nucleic acid and changes the accessibility of bases to cleavage reagents, the observed cleavage pattern will change. This method can be used to identify 10 and characterize the binding of small molecules to RNA as described, for example, by Prudent *et al.*, 1995, J. Am. Chem. Soc. 117:10145-10146 and Mei *et al.*, 1998, Biochem. 37:14204-14212.

15 In the preferred embodiment of this technique, the detectably labeled target nucleic acid is incubated with an individual test compound and then subjected to treatment with a cleavage reagent, either enzymatic or chemical. The reaction mixture can be 20 preferably be examined directly, or treated further to isolate and concentrate the nucleic acid. The fragments produced are separated by electrophoresis and the pattern of cleavage can be compared to a cleavage reaction performed in the absence of test compound. A change in the cleavage pattern directly indicates that the test compound binds to the target nucleic acid. Multiple test compounds can be examined both in parallel and serially.

25 Other embodiments of electrophoretic separation include, but are not limited to urea gel electrophoresis, gel filtration, pulsed field gel electrophoresis, two dimensional gel electrophoresis, continuous flow electrophoresis, zone electrophoresis, and agarose gel electrophoresis.

25

5.5.2. Fluorescence Spectroscopy

In a preferred embodiment, fluorescence polarization spectroscopy, an optical detection method that can differentiate the proportion of a fluorescent molecule that is either bound or unbound in solution (*e.g.*, the labeled target nucleic acid of the present invention), can be used to read reaction results without electrophoretic separation of the samples. Fluorescence polarization spectroscopy can be used to read the reaction results in the chip system disclosed in U.S. Patent Nos. 5,699,157; 5,842,787; 5,869,004; 5,876,675; 30 5,942,443; 5,948,227; 6,042,709; 6,042,710; 6,046,056; 6,048,498; 6,086,740; 6,132,685; 6,150,119; 6,150,180; 6,153,073; 6,167,910; 6,171,850; and 6,186,660, the disclosures of which are incorporated by reference in their entireties. The application of fluorescence 35

polarization spectroscopy to the chip system disclosed in the U.S. Patents listed *supra* is fast, efficient, and well-adapted for high-throughput screening.

In another embodiment, a compound that has an affinity for the target nucleic acid of interest can be labeled with a fluorophore to screen for test compounds that bind to the target nucleic acid. For example, a pyrene-containing aminoglycoside analog was used to accurately monitor antagonist binding to a prokaryotic 16S rRNA A site (which comprises the natural target for aminoglycoside antibiotics) in a screen using a fluorescence quenching technique in a 96-well plate format (Hamasaki & Rando, 1998, *Anal. Biochem.* 261(2):183-90).

In another embodiment, fluorescence resonance energy transfer (FRET) can be used to screen for test compounds that bind to the target nucleic acid. FRET, a characteristic change in fluorescence, occurs when two fluorophores with overlapping emission and excitation wavelength bands are held together in close proximity, such as by a binding event. In the preferred embodiment, the fluorophore on the target nucleic acid and the fluorophore on the test compounds will have overlapping excitation and emission spectra such that one fluorophore (the donor) transfers its emission energy to excite the other fluorophore (the acceptor). The acceptor preferably emits light of a different wavelength upon relaxing to the ground state, or relaxes non-radiatively to quench fluorescence. FRET is very sensitive to the distance between the two fluorophores, and allows measurement of molecular distances less than 10 nm. For example, U.S. Patent 6,337,183 to Arenas *et al.*, which is incorporated by reference in its entirety, describes a screen for compounds that bind RNA that uses FRET to measure the effect of test compounds on the stability of a target RNA molecule where the target RNA is labeled with both fluorescent acceptor and donor molecules and the distance between the two fluorophores as determined by FRET provides a measure of the folded structure of the RNA. Matsumoto *et al.* (2000, *Bioorg. Med. Chem. Lett.* 10:1857-1861) describe a system where a peptide that binds to HIV-1 TAR RNA is labeled on one end with a fluorescein fluorophore and a tetramethylrhodamine on the other end. The conformational change of the peptide upon binding to the RNA provided a FRET signal to screen for compounds that bound to the TAR RNA.

In the preferred embodiment, both the target nucleic acid and a compound that has an affinity for the target nucleic acid of interest are labeled with fluorophores with overlapping emission and excitation spectra (donor and acceptor), including but not limited to fluorescein and derivatives, rhodamine and derivatives, cyanine dyes and derivatives, bora-3a,4a-diaza-s-indacene (BODIPY®) and derivatives, pyrene, nanoparticles, or

non-fluorescent quenching molecules. Binding of a labeled test compound to the target nucleic acid can be identified by the change in observable fluorescence as a result of FRET.

If the target nucleic acid is labeled with the donor fluorophore, then the test compounds is labeled with the acceptor fluorophore. Conversely, if the target nucleic acid is labeled with the acceptor fluorophore, then the test compounds is labeled with the donor fluorophore. A wide variety of labels may be used, with the choice of label depending on sensitivity required, ease of conjugation with the compound, stability requirements, available instrumentation, and disposal provisions. The fluorophore on the target nucleic acid must be in close proximity to the binding site of the test compounds, but should not be incorporated into a target nucleic acid at the specific binding site at which test compounds are likely to bind, since the presence of a covalently attached label might interfere sterically or chemically with the binding of the test compounds at this site.

In yet another embodiment, homogeneous time-resolved fluorescence (“HTRF”) techniques based on time-resolved energy transfer from lanthanide ion complexes to a suitable acceptor species can be adapted for high-throughput screening for inhibitors of RNA-protein complexes (Hemmilä, 1999, J. Biomol. Screening 4:303-307; Mathis, 1999, J. Biomol. Screening 4:309-313). HTRF is similar to fluorescence resonance energy transfer using conventional organic dye pairs, but has several advantages, such as increased sensitivity and efficiency, and background elimination (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356).

Fluorescence spectroscopy has traditionally been used to characterize DNA-protein and protein-protein interactions, but fluorescence spectroscopy has not been widely used to characterize RNA-protein interactions because of an interfering absorption of RNA nucleotides with the intrinsic tryptophan fluorescence of proteins (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356.). However, fluorescence spectroscopy has been used in studying the single tryptophan residue within the arginine-rich RNA-binding domain of Rev protein and its interaction with the RRE in a time-resolved fluorescence study (Kwon & Carson, 1998, Anal. Biochem. 264:133-140). Thus, in this invention, fluorescence spectroscopy is less preferred if the test compounds or peptides or proteins possess intrinsic tryptophan fluorescence. However, fluorescence spectroscopy can be used for test compounds that do not possess intrinsic fluorescence.

5.5.3. Surface Plasmon Resonance (“SPR”)

Surface plasmon resonance (SPR) can be used for determining kinetic rate constants and equilibrium constants for macromolecular interactions by following the

association project in "real time" (Schuck, 1997, Annu. Rev. Biophys. Biomol. Struct. 26:541-566).

The principle of SPR is summarized by Xavier *et al.* (Trends Biotechnol., 2000, 18(8):349-356) as follows. Total internal reflection occurs at the boundary between two substances of different refractive index. The incident light's electromagnetic field penetrates beyond the interface as an evanescent wave, which extends a few hundred nanometers beyond the surface into the medium. Insertion of a thin gold foil at the interface produced SPR owing to the absorption of the energy from the evanescent wave by free electron clouds of the metal (plasmons). As a result of this absorbance, there is a drop in the intensity of the reflected light at a particular angle of incidence. The evanescent wave profile depends exquisitely on the refractive index of the medium it probes. Thus, the angle at which absorption occurs is very sensitive to the refractive changes in the external medium. All proteins and nucleic acids are known to change the refractive index of water by a similar amount per unit mass, irrespective of their amino acid or nucleotide composition (the refractive index change is different for proteins and nucleic acids). When the protein or nucleic acid content of the layer at the sensor changes, the refractive index also changes. Typically, one member of a complex is immobilized in a dextran layer and then the other member is introduced into the solution, either in a flow cell (Biacore AB, Uppsala, Sweden) or a stirred cuvette (Affinity Sensors, Santa Fe, New Mexico). It has been determined that there is a linear correlation between the surface concentration of protein or nucleic acid and the shift in resonance angle, which can be used to quantitate kinetic rate constants and/or the equilibrium constants.

In the present invention, the target RNA may be immobilized to the sensor surface through a streptavidin-biotin linkage, the linkage of which is disclosed by Crouch *et al.* (Methods Mol. Biol., 1999, 118:143-160). The RNA is biotinylated either during synthesis or post-synthetically via the conversion of the 3' terminal ribonucleoside of the RNA into a reactive free amino group or using a T7 polymerase incorporated guanosine monophosphorothioate at the 5' end. SPR has been used to determine the stoichiometry and affinity of the interaction between the HIV Rev protein and the RRE (Van Ryk & Venkatesan, 1999, J. Biol. Chem. 274:17452-17463) and the aminoglycoside antibiotics with RRE and a model RNA derived from the 16S ribosomal A site, respectively (Hendrix *et al.*, 1997, J. Am. Chem. Soc. 119:3641-3648; Wong *et al.*, 1998, Chem. Biol. 5:397-406).

In one embodiment of the present invention, the target nucleic acid can be immobilized to a sensor surface (*e.g.*, by a streptavidin-biotin linkage) and SPR can be used

to (a) determine whether the target RNA binds a test compound and (b) further characterize the binding of the target nucleic acids of the present invention to a test compound.

5 5.5.4. Mass Spectrometry

An automated method for analyzing mass spectrometer data which can analyze complex mixtures containing many thousands of components and can correct for background noise, multiply charged peaks and atomic isotope peaks is described in U.S. Patent No. 6,147,344, which is hereby incorporated by reference in its entirety. The system disclosed in U.S. Patent No. 6,147,344 is a method for analyzing mass spectrometer data in which a control sample measurement is performed providing a background noise check. The peak height and width values at each m/z ratio as a function of time are stored in a memory. A mass spectrometer operation on a material to be analyzed is performed and the peak height and width values at each m/z ratio versus time are stored in a second memory location. The mass spectrometer operation on the material to be analyzed is repeated a fixed number of times and the stored control sample values at each m/z ratio level at each time increment are subtracted from each corresponding one from the operational runs, thus producing a difference value at each mass ratio for each of the multiple runs at each time increment. If the MS value minus the background noise does not exceed a preset value, the m/z ratio data point is not recorded, thus eliminating background noise, chemical noise and false positive peaks from the mass spectrometer data. The stored data for each of the multiple runs is then compared to a predetermined value at each m/z ratio and the resultant series of peaks, which are now determined to be above the background, is stored in the m/z points in which the peaks are of significance.

25 One possibility for the utilization of mass spectrometry in high throughput screening is the integration of SPR with mass spectrometry. Approaches that have been tried are direct analysis of the analyte retained on the sensor chip and mass spectrometry with the eluted analyte (Sonksen *et al.*, 1998, Anal. Chem. 70:2731-2736; Nelson & Krone, 1999, J. Mol. Recog. 12:77-93). Further developments, especially in the interfacing of the sensor chip with the mass spectrometer and in reusing the sensor chip, are required to make 30 SPR combined with mass spectroscopy a high-throughput method for biomolecular interaction analysis and the screening of targets for small molecule inhibitors (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356).

35 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by any of the mass spectrometry processed described *supra*. Furthermore, mass spectrometry can also be used to elucidate

the structure of the test compound.

5.5.5. Scintillation Proximity Assay (“SPA”)

Scintillation Proximity Assay (“SPA”) is a method that can be used for screening small molecules that bind to the target RNAs. SPA would involve radiolabeling either the target RNA or the test compound and then quantitating its binding to the other member to a bead or a surface impregnated with a scintillant (Cook, 1996, Drug Discov. Today 1:287-294). Currently, fluorescence-based techniques are preferred for high-throughput screening (Pope *et al.*, 1999, Drug Discov. Today 4:350-362).

Screening for small molecules that inhibit Tat peptide:TAR RNA interaction has been performed with SPA, and inhibitors of the interaction were isolated and characterized (Mei *et al.*, 1997, Bioorg. Med. Chem. 5:1173-1184; Mei *et al.*, 1998, Biochemistry 37:14204-14212). A similar approach can be used to identify small molecules that directly bind to a preselected target RNA element in accordance with the invention can be utilized.

SPA can be adapted to high throughput screening by the availability of microplates, wherein the scintillant is directly incorporated into the plastic of the microtiter wells (Nakayama *et al.*, 1998, J. Biomol. Screening 3:43-48). Thus, one embodiment of the present invention comprises (a) labeling of the target nucleic acid with a radioactive or fluorescent label; (b) contacted the labeled nucleic acid with test compounds, wherein each test compound is in a microtiter well coated with scintillant and is tethered to the microtiter well; and (c) identifying and quantifying the test compounds bound to the target nucleic acid with SPA, wherein the test compound is identified by virtue of its location in the microplate.

5.5.6. Structure-Activity Relationships (“SAR”) by NMR Spectroscopy

NMR spectroscopy is a valuable technique for identifying complexed target nucleic acids by qualitatively determining changes in chemical shift, specifically from distances measured using relaxation effects, and NMR-based approaches have been used in the identification of small molecule binders of protein drug targets (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). The determination of structure-activity relationships (“SAR”) by NMR is the first method for NMR described in which small molecules that bind adjacent subsites are identified by two-dimentional ^1H - ^{15}N spectra of the target protein (Shuker *et al.*, 1996, Science 274:1531-1534). The signal from the bound molecule is monitored by employing line broadening, transferred NOEs and pulsed field gradient

diffusion measurements (Moore, 1999, Curr. Opin. Biotechnol. 10:54-58). A strategy for lead generation by NMR using a library of small molecules has been recently described (Fejzo *et al.*, 1999, Chem. Biol. 6:755-769).

5 In one embodiment of the present invention, the target nucleic acid complexed to a test compound can be determined by SAR by NMR. Furthermore, SAR by NMR can also be used to elucidate the structure of the test compound.

5.5.7. Size Exclusion Chromatography

10 In another embodiment of the present invention, size-exclusion chromatography is used to purify test compounds that are bound to a target nucleic acid from a complex mixture of compounds. Size-exclusion chromatography separates molecules based on their size and uses gel-based media comprised of beads with specific size distributions. When applied to a column, this media settles into a tightly packed matrix and forms a complex array of pores. Separation is accomplished by the inclusion or exclusion 15 of molecules by these pores based on molecular size. Small molecules are included into the pores and, consequently, their migration through the matrix is retarded due to the added distance they must travel before elution. Large molecules are excluded from the pores and migrate with the void volume when applied to the matrix. In the present invention, a target nucleic acid is incubated with a mixture of test compounds while free in solution and 20 allowed to reach equilibrium. When applied to a size exclusion column, test compounds free in solution are retained by the column, and test compounds bound to the target nucleic acid are passed through the column. In a preferred embodiment, spin columns commonly used for "desalting" of nucleic acids will be employed to separate bound from unbound test 25 compounds (*e.g.*, Bio-Spin columns manufactured by BIO-RAD). In another embodiment, the size exclusion matrix is packed into multiwell plates to allow high throughput separation of mixtures (*e.g.*, PLASMID 96-well SEC plates manufactured by Millipore).

5.5.8. Affinity Chromatography

30 In one embodiment of the present invention, affinity capture is used to purify test compounds that are bound to a target nucleic acid labeled with an affinity tag from a complex mixture of compounds. To accomplish this, a target nucleic acid labeled with an affinity tag is incubated with a mixture of test compounds while free in solution and then captured to a solid support once equilibrium has been established; alternatively, target 35 nucleic acids labeled with an affinity tag can be captured to a solid support first and then allowed to reach equilibrium with a mixture of test compounds.

The solid support is typically comprised of, but not limited to, cross-linked agarose beads that are coupled with a ligand for the affinity tag. Alternatively, the solid support may be a glass, silicon, metal, or carbon, plastic (polystyrene, polypropylene) surface with or without a self-assembled monolayer (SAM) either with a covalently attached ligand for the affinity tag, or with inherent affinity for the tag on the target nucleic acid.

Once the complex between the target nucleic acid and test compound has reached equilibrium and has been captured, one skilled in the art will appreciate that the retention of bound compounds and removal of unbound compounds is facilitated by washing the solid support with large excesses of binding reaction buffer. Furthermore, retention of high affinity compounds and removal of low affinity compounds can be accomplished by a number of means that increase the stringency of washing; these means include, but are not limited to, increasing the number and duration of washes, raising the salt concentration of the wash buffer, addition of detergent or surfactant to the wash buffer, and addition of non-specific competitor to the wash buffer.

In one embodiment, the test compounds themselves are detectably labeled with fluorescent dyes, radioactive isotopes, or nanoparticles. When the test compounds are applied to the captured target nucleic acid in a spatially addressed fashion (*e.g.*, in separate wells of a 96-well microplate), binding between the test compounds and the target nucleic acid can be determined by the presence of the detectable label on the test compound using fluorescence.

Following the removal of unbound compounds, bound compounds with high affinity for the target nucleic acid can be eluted from the immobilized target nucleic acids and analyzed. The elution of test compounds can be accomplished by any means that break the non-covalent interactions between the target nucleic acid and compound. Means for elution include, but are not limited to, changing the pH, changing the salt concentration, the application of organic solvents, and the application of molecules that compete with the bound ligand. In a preferred embodiment, the means employed for elution will release the compound from the target RNA, but will not effect the interaction between the affinity tag and the solid support, thereby achieving selective elution of test compound. Moreover, a preferred embodiment will employ an elution buffer that is volatile to allow for subsequent concentration by lyophilization of the eluted compound (*e.g.*, 0 M to 5 M ammonium acetate).

5.5.9. Nanoparticle Aggregation

In one embodiment of the present invention, both the target nucleic acid and the test compounds are labeled with nanoparticles. A nanoparticle is a cluster of ions with controlled size from 0.1 to 1000 nm comprised of metals, metal oxides, or semiconductors including, but not limited to Ag_2S , ZnS , CdS , CdTe , Au , or TiO_2 . Methods for the attachment of nucleic acids and small molecules to nanoparticles are well known to one of skill in the art (reviewed in Niemeyer, 2001, Angew. Chem. Int. Ed. 40:4129-4158. The references cited therein are hereby incorporated by reference in their entireties). In particular, if multiple copies of the target nucleic acid are attached to a single nanoparticle and multiple copies of a test compound are attached to another nanoparticle, then interaction between the test compound and target nucleic acid will induce aggregation of nanoparticles as described, for example, by Mitchel *et al.* 1999, J. Am. Chem. Soc. 121:8122-8123. The aggregate can be detected by changes in absorbance or fluorescence spectra and physically separated from the unbound components through filtration or centrifugation.

5.6. Methods for Identifying or Characterizing the Test Compounds Bound to the Target Nucleic Acids

If the library comprises arrays or microarrays of test compounds, wherein each test compound has an address or identifier, the test compound can be deconvoluted, e.g., by cross-referencing the positive sample to original compound list that was applied to the individual test assays.

If the library is a peptide or nucleic acid library, the sequence of the test compound can be determined by direct sequencing of the peptide or nucleic acid. Such methods are well known to one of skill in the art.

A number of physico-chemical techniques can be used for the de novo characterization of test compounds bound to the target.

5.6.1. Mass Spectrometry

Mass spectrometry (e.g., electrospray ionization ("ESI") and matrix-assisted laser desorption-ionization ("MALDI"), Fourier-transform ion cyclotron resonance ("FT-ICR")) can be used both for high-throughput screening of test compounds that bind to a target RNA and elucidating the structure of the test compound. Thus, one example of mass spectroscopy is that separation of a bound and unbound complex and test compound structure elucidation can be carried out in a single step.

MALDI uses a pulsed laser for desorption of the ions and a time-of-flight analyzer, and has been used for the detection of noncovalent tRNA:amino-acyl-tRNA synthetase complexes (Gruic-Sovulj *et al.*, 1997, J. Biol. Chem. 272:32084-32091).
5 However, covalent cross-linking between the target nucleic acid and the test compound is required for detection, since a non-covalently bound complex may dissociate during the MALDI process.

ESI mass spectrometry ("ESI-MS") has been of greater utility for studying non-covalent molecular interactions because, unlike the MALDI process, ESI-MS generates molecular ions with little to no fragmentation (Xavier *et al.*, 2000, Trends Biotechnol. 10 18(8):349-356). ESI-MS has been used to study the complexes formed by HIV Tat peptide and protein with the TAR RNA (Sannes-Lowery *et al.*, 1997, Anal. Chem. 69:5130-5135).

Fourier-transform ion cyclotron resonance ("FT-ICR") mass spectrometry provides high-resolution spectra, isotope-resolved precursor ion selection, and accurate mass assignments (Xavier *et al.*, 2000, Trends Biotechnol. 18(8):349-356). FT-ICR has been used to study the interaction of aminoglycoside antibiotics with cognate and non-cognate RNAs (Hofstadler *et al.*, 1999, Anal. Chem. 71:3436-3440; Griffey *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:10129-10133). As true for all of the mass spectrometry methods discussed herein, FT-ICR does not require labeling of the target RNA or a test compound.
20

An advantage of mass spectroscopy is not only the elucidation of the structure of the test compound, but also the determination of the structure of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.
25

5.6.2. NMR Spectroscopy

As described above, NMR spectroscopy is a technique for identifying binding sites in target nucleic acids by qualitatively determining changes in chemical shift, 30 specifically from distances measured using relaxation effects. Examples of NMR that can be used for the invention include, but are not limited to, one-dimentional NMR, two-dimentional NMR, correlation spectroscopy ("COSY"), and nuclear Overhauser effect ("NOE") spectroscopy. Such methods of structure determination of test compounds are well known to one of skill in the art.

Similar to mass spectroscopy, an advantage of NMR is the not only the elucidation of the structure of the test compound, but also the determination of the structure
35

of the test compound bound to the preselected target RNA. Such information can enable the discovery of a consensus structure of a test compound that specifically binds to a preselected target RNA.

5

5.6.3. Vibrational Spectroscopy

Vibrational spectroscopy (e.g. infrared (IR) spectroscopy or Raman spectroscopy) can be used for elucidating the structure of the test compound on the isolated bead.

10

Infrared spectroscopy measures the frequencies of infrared light (wavelengths from 100 to 10,000 nm) absorbed by the test compound as a result of excitation of vibrational modes according to quantum mechanical selection rules which require that absorption of light cause a change in the electric dipole moment of the molecule. The infrared spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

15

Infrared spectra can be measured in a scanning mode by measuring the absorption of individual frequencies of light, produced by a grating which separates frequencies from a mixed-frequency infrared light source, by the test compound relative to a standard intensity (double-beam instrument) or pre-measured ('blank') intensity (single-beam instrument). In a preferred embodiment, infrared spectra are measured in a pulsed mode (FT-IR) where a mixed beam, produced by an interferometer, of all infrared light frequencies is passed through or reflected off the test compound. The resulting interferogram, which may or may not be added with the resulting interferograms from subsequent pulses to increase the signal strength while averaging random noise in the electronic signal, is mathematically transformed into a spectrum using Fourier Transform or Fast Fourier Transform algorithms.

20

25

30

Raman spectroscopy measures the difference in frequency due to absorption of infrared frequencies of scattered visible or ultraviolet light relative to the incident beam. The incident monochromatic light beam, usually a single laser frequency, is not truly absorbed by the test compound but interacts with the electric field transiently. Most of the light scattered off the sample will be unchanged (Rayleigh scattering) but a portion of the scatter light will have frequencies that are the sum or difference of the incident and molecular vibrational frequencies. The selection rules for Raman (inelastic) scattering require a change in polarizability of the molecule. While some vibrational transitions are observable in both infrared and Raman spectrometry, most are observable only with one or

35

the other technique. The Raman spectrum of any molecule is a unique pattern of absorption wavelengths of varying intensity that can be considered as a molecular fingerprint to identify any compound.

5 Raman spectra are measured by submitting monochromatic light to the sample, either passed through or preferably reflected off, filtering the Rayleigh scattered light, and detecting the frequency of the Raman scattered light. An improved Raman spectrometer is described in US Patent No. 5,786,893 to Fink *et al.*, which is hereby incorporated by reference.

10 Vibrational microscopy can be measured in a spatially resolved fashion to address single beads by integration of a visible microscope and spectrometer. A microscopic infrared spectrometer is described in U.S. Patent No. 5,581,085 to Reffner *et al.*, which is hereby incorporated by reference in its entirety. An instrument that simultaneously performs a microscopic infrared and microscopic Raman analysis on a 15 sample is described in U.S. Patent No. 5,841,139 to Sostek *et al.*, which is hereby incorporated by reference in its entirety.

20 In the preferred embodiment, test compounds can be identified by matching the IR or Raman spectra of a test compound to a dataset of vibrational (IR or Raman) spectra previously acquired for each compound in the combinatorial library. By this method, the spectra of compounds with known structure are recorded so that comparison with these spectra can identify compounds again when isolated from RNA binding experiments.

5.7. Secondary Biological Screens

25 The test compounds identified in the binding assay (for convenience referred to herein as a "lead" compound) can be tested for biological activity using host cells containing or engineered to contain the target RNA element coupled to a functional readout system. For example, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene. In this example, the 30 lead compounds are assayed in the presence or absence of the target RNA. Alternatively, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound.

35 In one embodiment, the lead compound can be tested in a host cell engineered to contain the target RNA element controlling the expression of a reporter gene, such as, but not limited to, β -galactosidase, green fluorescent protein, red fluorescent protein, luciferase, chloramphenicol acetyltransferase, alkaline phosphatase, and β -

lactamase. In a preferred embodiment, a cDNA encoding the target element is fused upstream to a reporter gene wherein translation of the reporter gene is repressed upon binding of the lead compound to the target RNA. In other words, the steric hindrance caused by the binding of the lead compound to the target RNA repressed the translation of the reporter gene. This method, termed the translational repression assay procedure ("TRAP") has been demonstrated in *E. coli* and *S. cerevisiae* (Jain & Belasco, 1996, Cell 87(1):115-25; Huang & Schreiber, 1997, Proc. Natl. Acad. Sci. USA 94:13396-13401).

In another embodiment, a phenotypic or physiological readout can be used to assess activity of the target RNA in the presence and absence of the lead compound. For example, the target RNA may be overexpressed in a cell in which the target RNA is endogenously expressed. Where the target RNA controls expression of a gene product involved in cell growth or viability, the *in vivo* effect of the lead compound can be assayed by measuring the cell growth or viability of the target cell. Alternatively, a reporter gene can also be fused downstream of the target RNA sequence and the effect of the lead compound on reporter gene expression can be assayed.

Alternatively, the lead compounds identified in the binding assay can be tested for biological activity using animal models for a disease, condition, or syndrome of interest. These include animals engineered to contain the target RNA element coupled to a functional readout system, such as a transgenic mouse. Animal model systems can also be used to demonstrate safety and efficacy.

Compounds displaying the desired biological activity can be considered to be lead compounds, and will be used in the design of congeners or analogs possessing useful pharmacological activity and physiological profiles. Following the identification of a lead compound, molecular modeling techniques can be employed, which have proven to be useful in conjunction with synthetic efforts, to design variants of the lead that can be more effective. These applications may include, but are not limited to, Pharmacophore Modeling (*cf.* Lamothe, *et al.* 1997, J. Med. Chem. 40: 3542; Mottola *et al.* 1996, J. Med. Chem. 39: 285; Beusen *et al.* 1995, Biopolymers 36: 181; P. Fossa *et al.* 1998, Comput. Aided Mol. Des. 12: 361), QSAR development (*cf.* Siddiqui *et al.* 1999, J. Med. Chem. 42: 4122; Barreca *et al.* 1999 Bioorg. Med. Chem. 7: 2283; Kroemer *et al.* 1995, J. Med. Chem. 38: 4917; Schaal *et al.* 2001, J. Med. Chem. 44: 155; Buolamwini & Assefa 2002, J. Mol. Chem. 45: 84), Virtual docking and screening/scoring (*cf.* Anzini *et al.* 2001, J. Med. Chem. 44: 1134; Faaland *et al.* 2000, Biochem. Cell. Biol. 78: 415; Silvestri *et al.* 2000, Bioorg. Med. Chem. 8: 2305; J. Lee *et al.* 2001, Bioorg. Med. Chem. 9: 19), and Structure Prediction using RNA structural programs including, but not limited to mFold (as described

by Zuker *et al.* Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide in RNA Biochemistry and Biotechnology pp. 11-43, J. Barciszewski & B.F.C. Clark, eds. (NATO ASI Series, Kluwer Academic Publishers, 1999) and Mathews *et al.* 1999 J. Mol. Biol. 288: 911-940); RNAMotif (Macke *et al.* 2001, Nucleic Acids Res. 29: 4724-4735; and the Vienna RNA package (Hofacker *et al.* 1994, Monatsh. Chem. 125: 167-188).

Further examples of the application of such techniques can be found in several review articles, such as Rotivinen *et al.*, 1988, Acta Pharmaceutical Fennica 97:159-166; Ripka, 1998, New Scientist 54-57; McKinlay & Rossmann, 1989, Annu. Rev. Pharmacol. Toxicol. 29:111-122; Perry & Davies, QSAR: Quantitative Structure-Activity Relationships in Drug Design pp. 189-193 (Alan R. Liss, Inc. 1989); Lewis & Dean, 1989, Proc. R. Soc. Lond. 236:125-140 and 141-162; Askew *et al.*, 1989, J. Am. Chem. Soc. 111:1082-1090. Molecular modeling tools employed may include those from Tripos, Inc., St. Louis, Missouri (e.g., Sybyl/UNITY, CONCORD, DiverseSolutions), Accelrys, San Diego, California (e.g., Catalyst, Wisconsin Package {BLAST, etc.}), Schrodinger, Portland, Oregon (e.g., QikProp, QikFit, Jaguar) or other such vendors as BioDesign, Inc. (Pasadena, California), Allelix, Inc. (Mississauga, Ontario, Canada), and Hypercube, Inc. (Cambridge, Ontario, Canada), and may include privately designed and/or "academic" software (e.g. RNAMotif, mFOLD). These application suites and programs include tools for the atomistic construction and analysis of structural models for drug-like molecules, proteins, and DNA or RNA and their potential interactions. They also provide for the calculation of important physical properties, such as solubility estimates, permeability metrics, and empirical measures of molecular "druggability" (e.g., Lipinski "Rule of 5" as described by Lipinski *et al.* 1997, Adv. Drug Delivery Rev. 23: 3-25). Most importantly, they provide appropriate metrics and statistical modeling power (such as the patented CoMFA technology in Sybyl as described in US Patents 6,240,374 and 6,185,506) to develop Quantitative Structural Activity Relationships (QSARs) which are used to guide the synthesis of more efficacious clinical development candidates while improving desirable physical properties, as determined by results from the aforementioned secondary screening protocols.

5.8. Use of Identified Compounds That Bind RNA to Treat/Prevent Disease

Biologically active compounds identified using the methods of the invention or a pharmaceutically acceptable salt thereof can be administered to a patient, preferably a mammal, more preferably a human, suffering from a disease whose progression is

associated with a target RNA:host cell factor interaction *in vivo*. In certain embodiments, such compounds or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*.

5 In one embodiment, "treatment" or "treating" refers to an amelioration of a disease, or at least one discernible symptom thereof. In another embodiment, "treatment" or "treating" refers to an amelioration of at least one measurable physical parameter, not necessarily discernible by the patient. In yet another embodiment, "treatment" or "treating" refers to inhibiting the progression of a disease, either physically, *e.g.*, stabilization of a
10 discernible symptom, physiologically, *e.g.*, stabilization of a physical parameter, or both. In yet another embodiment, "treatment" or "treating" refers to delaying the onset of a disease.

15 In certain embodiments, the compound or a pharmaceutically acceptable salt thereof is administered to a patient, preferably a mammal, more preferably a human, as a preventative measure against a disease associated with an RNA:host cell factor interaction *in vivo*. As used herein, "prevention" or "preventing" refers to a reduction of the risk of acquiring a disease. In one embodiment, the compound or a pharmaceutically acceptable salt thereof is administered as a preventative measure to a patient. According to this embodiment, the patient can have a genetic predisposition to a disease, such as a family history of the disease, or a non-genetic predisposition to the disease. Accordingly, the
20 compound and pharmaceutically acceptable salts thereof can be used for the treatment of one manifestation of a disease and prevention of another.

25 When administered to a patient, the compound or a pharmaceutically acceptable salt thereof is preferably administered as component of a composition that optionally comprises a pharmaceutically acceptable vehicle. The composition can be administered orally, or by any other convenient route, for example, by infusion or bolus injection, by absorption through epithelial or mucocutaneous linings (*e.g.*, oral mucosa, rectal, and intestinal mucosa, *etc.*) and may be administered together with another biologically active agent. Administration can be systemic or local. Various delivery systems are known, *e.g.*, encapsulation in liposomes, microparticles, microcapsules, capsules, *etc.*, and can be used to administer the compound and pharmaceutically acceptable salts thereof.

30 Methods of administration include but are not limited to intradermal, intramuscular, intraperitoneal, intravenous, subcutaneous, intranasal, epidural, oral, sublingual, intranasal, intracerebral, intravaginal, transdermal, rectally, by inhalation, or topically, particularly to the ears, nose, eyes, or skin. The mode of administration is left to
35

the discretion of the practitioner. In most instances, administration will result in the release of the compound or a pharmaceutically acceptable salt thereof into the bloodstream.

5 In specific embodiments, it may be desirable to administer the compound or a pharmaceutically acceptable salt thereof locally. This may be achieved, for example, and not by way of limitation, by local infusion during surgery, topical application, e.g., in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers.

10 In certain embodiments, it may be desirable to introduce the compound or a pharmaceutically acceptable salt thereof into the central nervous system by any suitable route, including intraventricular, intrathecal and epidural injection. Intraventricular injection may be facilitated by an intraventricular catheter, for example, attached to a reservoir, such as an Ommaya reservoir.

15 Pulmonary administration can also be employed, e.g., by use of an inhaler or nebulizer, and formulation with an aerosolizing agent, or via perfusion in a fluorocarbon or synthetic pulmonary surfactant. In certain embodiments, the compound and pharmaceutically acceptable salts thereof can be formulated as a suppository, with traditional binders and vehicles such as triglycerides.

20 In another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a vesicle, in particular a liposome (see Langer, 1990, Science 249:1527-1533; Treat *et al.*, in *Liposomes in the Therapy of Infectious Disease and Cancer*, Lopez-Berestein and Fidler (eds.), Liss, New York, pp. 353-365 (1989); Lopez-Berestein, *ibid.*, pp. 317-327; see generally *ibid.*).

25 In yet another embodiment, the compound and pharmaceutically acceptable salts thereof can be delivered in a controlled release system (see, e.g., Goodson, in *Medical Applications of Controlled Release*, *supra*, vol. 2, pp. 115-138 (1984)). Other controlled-release systems discussed in the review by Langer, 1990, Science 249:1527-1533) may be used. In one embodiment, a pump may be used (see Langer, *supra*; Sefton, 1987, *CRC Crit. Ref. Biomed. Eng.* 14:201; Buchwald *et al.*, 1980, *Surgery* 88:507 Saudek *et al.*, 1989, *N. Engl. J. Med.* 321:574). In another embodiment, polymeric materials can be used (see *Medical Applications of Controlled Release*, Langer and Wise (eds.), CRC Pres., Boca Raton, Florida (1974); *Controlled Drug Bioavailability, Drug Product Design and Performance*, Smolen and Ball (eds.), Wiley, New York (1984); Ranger and Peppas, 1983, *J. Macromol. Sci. Rev. Macromol. Chem.* 23:61; see also Levy *et al.*, 1985, *Science*

228:190; During *et al.*, 1989, Ann. Neurol. 25:351; Howard *et al.*, 1989, J. Neurosurg. 71:105). In yet another embodiment, a controlled-release system can be placed in proximity of a target RNA of the compound or a pharmaceutically acceptable salt thereof, thus requiring only a fraction of the systemic dose.

5 Compositions comprising the compound or a pharmaceutically acceptable salt thereof ("compound compositions") can additionally comprise a suitable amount of a pharmaceutically acceptable vehicle so as to provide the form for proper administration to the patient.

10 In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, mammals, and more particularly in humans. The term "vehicle" refers to a diluent, adjuvant, excipient, or carrier with which a compound of the invention is administered. Such pharmaceutical vehicles can be liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and the like. The pharmaceutical vehicles can be saline, gum acacia, gelatin, starch paste, talc, keratin, colloidal silica, urea, and the like. In addition, auxiliary, stabilizing, thickening, lubricating and coloring agents may be used. When administered to a patient, the pharmaceutically acceptable vehicles are preferably sterile. Water is a preferred vehicle when the compound of the invention is administered intravenously. Saline solutions and aqueous dextrose and glycerol solutions can also be employed as liquid vehicles, particularly for injectable solutions. Suitable pharmaceutical vehicles also include excipients such as starch, glucose, lactose, sucrose, gelatin, malt, rice, flour, chalk, silica gel, sodium stearate, glycerol monostearate, talc, sodium chloride, dried skim milk, glycerol, propylene, glycol, water, ethanol and the like. Compound compositions, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents.

15 Compound compositions can take the form of solutions, suspensions, emulsion, tablets, pills, pellets, capsules, capsules containing liquids, powders, sustained-release formulations, suppositories, emulsions, aerosols, sprays, suspensions, or any other form suitable for use. In one embodiment, the pharmaceutically acceptable vehicle is a capsule (see e.g., U.S. Patent No. 5,698,155). Other examples of suitable pharmaceutical vehicles are described in Remington's Pharmaceutical Sciences, Alfonso R. Gennaro, ed., Mack Publishing Co. Easton, PA, 19th ed., 1995, pp. 1447 to 1676, incorporated herein by reference.

In a preferred embodiment, the compound or a pharmaceutically acceptable salt thereof is formulated in accordance with routine procedures as a pharmaceutical composition adapted for oral administration to human beings. Compositions for oral delivery may be in the form of tablets, lozenges, aqueous or oily suspensions, granules, 5 powders, emulsions, capsules, syrups, or elixirs, for example. Orally administered compositions may contain one or more agents, for example, sweetening agents such as fructose, aspartame or saccharin; flavoring agents such as peppermint, oil of wintergreen, or cherry; coloring agents; and preserving agents, to provide a pharmaceutically palatable 10 preparation. Moreover, where in tablet or pill form, the compositions can be coated to delay disintegration and absorption in the gastrointestinal tract thereby providing a sustained action over an extended period of time. Selectively permeable membranes surrounding an osmotically active driving compound are also suitable for orally 15 administered compositions. In these later platforms, fluid from the environment surrounding the capsule is imbibed by the driving compound, which swells to displace the agent or agent composition through an aperture. These delivery platforms can provide an essentially zero order delivery profile as opposed to the spiked profiles of immediate release formulations. A time delay material such as glycerol monostearate or glycerol stearate may also be used. Oral compositions can include standard vehicles such as 20 mannitol, lactose, starch, magnesium stearate, sodium saccharine, cellulose, magnesium carbonate, and the like. Such vehicles are preferably of pharmaceutical grade. Typically, compositions for intravenous administration comprise sterile isotonic aqueous buffer. Where necessary, the compositions may also include a solubilizing agent.

In another embodiment, the compound or a pharmaceutically acceptable salt thereof can be formulated for intravenous administration. Compositions for intravenous 25 administration may optionally include a local anesthetic such as lignocaine to lessen pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water-free concentrate in a hermetically sealed container such as an ampoule or sachette indicating the 30 quantity of active agent. Where the compound or a pharmaceutically acceptable salt thereof is to be administered by infusion, it can be dispensed, for example, with an infusion bottle containing sterile pharmaceutical grade water or saline. Where the compound or a pharmaceutically acceptable salt thereof is administered by injection, an ampoule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to 35 administration.

The amount of a compound or a pharmaceutically acceptable salt thereof that will be effective in the treatment of a particular disease will depend on the nature of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed will also depend on the route of administration, and the seriousness of the disease, and should be decided according to the judgment of the practitioner and each patient's circumstances. However, suitable dosage ranges for oral administration are generally about 0.001 milligram to about 200 milligrams of a compound or a pharmaceutically acceptable salt thereof per kilogram body weight per day. In specific preferred embodiments of the invention, the oral dose is about 0.01 milligram to about 100 milligrams per kilogram body weight per day, more preferably about 0.1 milligram to about 75 milligrams per kilogram body weight per day, more preferably about 0.5 milligram to 5 milligrams per kilogram body weight per day. The dosage amounts described herein refer to total amounts administered; that is, if more than one compound is administered, or if a compound is administered with a therapeutic agent, then the preferred dosages correspond to the total amount administered. Oral compositions preferably contain about 10% to about 95% active ingredient by weight.

Suitable dosage ranges for intravenous (i.v.) administration are about 0.01 milligram to about 100 milligrams per kilogram body weight per day, about 0.1 milligram to about 35 milligrams per kilogram body weight per day, and about 1 milligram to about 10 milligrams per kilogram body weight per day. Suitable dosage ranges for intranasal administration are generally about 0.01 pg/kg body weight per day to about 1 mg/kg body weight per day. Suppositories generally contain about 0.01 milligram to about 50 milligrams of a compound of the invention per kilogram body weight per day and comprise active ingredient in the range of about 0.5% to about 10% by weight.

Recommended dosages for intradermal, intramuscular, intraperitoneal, subcutaneous, epidural, sublingual, intracerebral, intravaginal, transdermal administration or administration by inhalation are in the range of about 0.001 milligram to about 200 milligrams per kilogram of body weight per day. Suitable doses for topical administration are in the range of about 0.001 milligram to about 1 milligram, depending on the area of administration. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems. Such animal models and systems are well known in the art.

The compound and pharmaceutically acceptable salts thereof are preferably assayed *in vitro* and *in vivo*, for the desired therapeutic or prophylactic activity, prior to use

in humans. For example, *in vitro* assays can be used to determine whether it is preferable to administer the compound, a pharmaceutically acceptable salt thereof, and/or another therapeutic agent. Animal model systems can be used to demonstrate safety and efficacy.

5 A variety of compounds can be used for treating or preventing diseases in mammals. Types of compounds include, but are not limited to, peptides, peptide analogs including peptides comprising non-natural amino acids, *e.g.*, D-amino acids, phosphorous analogs of amino acids, such as α -amino phosphonic acids and α -amino phosphinic acids, or amino acids having non-peptide linkages, nucleic acids, nucleic acid analogs such as 10 phosphorothioates or peptide nucleic acids ("PNAs"), hormones, antigens, synthetic or naturally occurring drugs, opiates, dopamine, serotonin, catecholamines, thrombin, acetylcholine, prostaglandins, organic molecules, pheromones, adenosine, sucrose, glucose, lactose and galactose.

15 **6. EXAMPLE: THERAPEUTIC TARGETS**

The therapeutic targets presented herein are by way of example, and the present invention is not to be limited by the targets described herein. The therapeutic targets presented herein as DNA sequences are understood by one of skill in the art that the sequences can be converted to RNA sequences.

20

6.1. Tumor Necrosis Factor Alpha ("TNF- α ")

GenBank Accession # X01394:

1 gcagaggacc agctaagagg gagagaagca actacagacc cccctgaaa acaaccctca
61 gacgccat cccctgacaa gctgccaggc aggttctttt ccttcacat actgaccac
121 ggctccaccc tctctccctt gaaaaaggaca ccatgagcac tgaaagcatg atccgggacg
181 tggagctggc cgaggaggcg ctccccaga agacaggggg gccccaggc tccaggcggt
241 gcttgttctt cagcccttc tccttcttga tggcggcagg cgccaccacg ctcttctgccc
301 tgctgcactt tggagtgtatc ggccccaga gggaaagatgtt ccccaaggac ctctcttaa
361 tcagcccttggccaggca gtcagatcat ttctcgaac cccgagtgtac aagcctgttag
421 cccatgttgtt agcaaaccctt caagctgagg ggcagctcca gtggctgaac cgccgggcca
481 atgccttcctt ggccaatggc gtggagctga gagataacca gctgggttg ccatcagagg
541 gcctgtacctt catctactcc caggcttctt tcaagggcca aggctcccc tccaccatg
601 tgctcttcac ccacaccatc agccgcattcg ccgttcttta ccagaccaag gtcaacccctt
661 tcttgccat caagagcccc ttggcaggagg agacccaga gggggctgag gccaaggccctt
721 ggtatgagcc catctatctg ggaggggtcttccagcttggaa gaaggggtgac cgactcagcg
781 ctgagatcaa tcggccggac tatctcgact ttggccaggc tactttggaa

841 tcattgccct gtgaggagga cgaacatcca accttccaa acgcctcccc tgcccaatc
 901 cctttattac cccctccctc agacaccctc aaccttctt ggctaaaaa gagaattggg
 961 ggcttagggt cggaacccaa gcttagaact ttaagcaaca agaccaccac ttcaaact
 1021 gggattcagg aatgtgtggc ctgcacagt aattgctggc aaccactaag aattcaaact
 1081 gggcctcca gaacticactg gggcctacag ctttgatccc tgacatctgg aatctggaga
 1141 ccagggagcc ttgggtctg gccagaatgc tgcaggactt gagaagacct cacctagaaa
 1201 ttgacacaag tggacccttag gccttcctct ctccagatgt ttccagactt ctttgagaca
 1261 cgagcccg ccctccccat ggagccagct ccctctattt atgttgtcac ttgtgattat
 1321 ttattattta ttattatttt atttattttac agatgaatgt atttatttgg gagaccgggg
 1381 tatccctgggg gacccaatgt aggagctgcc ttggctcaga catgtttcc gtgaaaacgg
 1441 agctgaacaa taggctgttc ccatgttagcc ccctggcctc tgccttc tttgattatg
 1501 tttttaaaaa tatttatctg attaagtgtt ctaaacaatg ctgatttggt gaccaactgt
 1561 cactcattgc tgagcctctg ctcccccagg gagttgtgtc tctaattgcc ctactattca
 1621 gtttttttttataaaatgggg ctt (SEQ ID NO: 6)

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 152
- (2) 3' Untranslated Region - nts 852 - 1643

Initial Specific Target Motif:
 Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' AUUUAUUUAUUUAUUUAUUUA 3' (SEQ ID NO: 1)

25 6.2. Granulocyte-macrophage Colony Stimulating Factor (“GM-CSF”)
 GenBank Accession # NM_000758:

1 gctggaggat gtggctgcag agcctgctgc tcttggcac tgcgtggcgc agcatctcg
 61 cacccgccccg ctgcggcagc cccagcacgc agccctggga gcatgtgaat gccatccagg
 121 agggccggcgc ttcctgaac ctgagtagag acactgtgc tgagatgaat gaaacagtag
 181 aagtcatctc agaaatgtt gacccctcagg agccgacctg cctacagacc cgcctggagc
 241 tgtacaagca gggcctgcgg ggcagcctca ccaagctaa gggcccttg accatgtatgg
 30 301 ccagccacta caagcagcac tgccctccaa cccggaaac ttccctgtca accccagacta
 361 tcaccttiga aagttcaaa gagaacctga aggacttct gcttgcaccccttgc
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaaactcag gatggatc ttggaggac caaggggtgg
 541 gcccacagcca tggtggagt ggcctggacc tgccctgggc cacactgacc ctgatacagg

601 catggcagaa gaatgggaat attttatact gacagaaaatc agtaatattt atatatttat
 661 attttaaaa tattttata ttatattt taagtgcata ttccatattt attcaagatg
 721 ttaccgtataatttata ttaaaaatat gctct (SEQ ID NO: 7)

5

GenBank Accession # XM_003751:

1 tctggaggat gtggctgcag agcgtctgc tcgtggcac tgcgtgc acgtctcg
 61 cacccgcccc ctcgccccagc cccagcacgc agccctggga gcatgtgaat gcccattcagg
 121 aggccccggcg tctcgtaac ctgagtagag acactgctgc tgagatgaat gaaacagtag
 181 aagtcatctc agaaatgttt gacccctcagg agccgacccctg cctacagacc cgcctggagg
 241 tgtacaagca gggcctgcgg ggcagcccta ccaagctcaa gggcccttg accatgtgg
 301 ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtca acccagacta
 361 tcaccttga aagttcaaa gagaacctga aggacttct gcttgtcatc ccctttgact
 421 gctgggagcc agtccaggag tgagaccggc cagatgaggc tggccaagcc ggggagctgc
 481 tctctcatga aacaagagct agaaactcag gatggtcata ttggagggac caaggggtgg
 541 gccacagcca tgggggagt ggcctggacc tgccctgggc cacactgacc ctgatacagg
 601 catggcagaa gaatgggaat attttatact gacagaaaatc agtaatattt atatatttat
 661 attttaaaa tattttata ttatattt taagtgcata ttccatattt attcaagatg
 721 ttaccgtataatttata ttaaaaatat gctct (SEQ ID NO: 8)

20

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 32
- (2) 3' Untranslated Region - nts 468 - 789

25

Initial Specific Target Motif:

Group I AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' AUUUAUUUUUUUUUUUUA 3' (SEQ ID NO: 1)

6.3. Interleukin 2 ("IL-2")

30

GenBank Accession # U25676:

35

1 atcactctct ttaatcaacta ctccatcaa cctcaactcc tgccacaatg tacaggatgc
 61 aactcctgtc ttgcattgca ctaattctg cacttgcac aaacagtgc cctacttcaa
 121 gttcgacaaa gaaaacaaag aaaacacagc tacaactgga gcatttactg ctggattac
 181 agatgtttttaatggaaatt aataattaca agaatccaa actcaccagg atgctcacat
 241 ttaagtttttaatggaaatt aataattaca agaatccaa actcaccagg atgctcacat
 301 aactcaaacc tctggaggaa gtgtcaattt tagctcaag caaaaactt cacttaagac

361 ccagggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa
 421 cattcatgtg tgaatatgca gatgagacag caaccattgt agaatttctg aacagatgga
 481 ttacctttg tcaaagcatc atctcaacac taacttgata attaagtgct tcccacttaa
 5 541 aacatatcag gccttctatt tatttattta aatatttaaa ttttatattt atttgtaat
 601 gtatggttgc tacctattgt aacttattt cttaatctta aaactataaa tatggatctt
 661 ttatgattct tttgttaagc ccttagggct ctaaaatggt ttaccttattt tatcccaaaa
 721 atatttatta ttatgttga tgtaaaatat agtaictatg tagattgggt agtaaaaacta
 781 ttaataaaat ttgataaata taaaaaaaaaa aaacaaaaaaaaa aaaaa (SEQ ID NO: 9)

10 General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 47
- (2) 3' Untranslated Region - nts 519- 825

15 Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

6.4. Interleukin 6 ("IL-6")

20 GenBank Accession # NM_000600:

1 ttctgccctc gagccccaccg ggaacgaaag agaagctcta tctgcctcc aggagccag
 61 ctatgaactc cttccaca agcgccctcg gtccagttgc cttccctcg gggctgcctcc
 121 tggtgttgc tgctgccctc cctgccccag tacccccagg agaagattcc aaagatgttag
 181 ccgccccaca cagacagcca ctcacctt cagaacgaat tgacaaacaa attcggtaca
 241 tcctcgacgg catctcagcc ctgagaaagg agacatgtaa caagagtaac atgtgtgaaa
 301 gcagcaaaga ggcactggca gaaaacaacc tgaacctcc aaagatggct gaaaaagatg
 361 gatgttcca atctggattc aatgaggaga ctgcctggt gaaaatcatc actggcttt
 421 tggagttga ggtataccta ggttacctcc agaacagatt tgagagtagt gaggaacaag
 481 ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaga
 541 atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc
 601 aggcacagaa ccagtggctg caggacatga caactcatct cattctgcgc agcttaagg
 661 agttcctgca gtccagccctg agggcttc ggcaaatgta gcatggcac ctcaatgt
 721 tgggtttaat gggcattcc tcttctggtc agaaacctgt ccactggca cagaacttat
 781 gttgttctct atggagaact aaaagtatga gctttaggac actattttaa ttattttaa
 841 ttatataata tttaaatatg tgaagctgag ttaattatg taagtcatat ttatatttt
 901 aagaagtacc acttgaaaca ttatgtat tagtttgaa ataataatgg aaagtggctt

961 tgcagttga atatccttg ttcagagcc agatcattc ttggaaagtg taggcttacc
 1021 tcaaataaat ggctaactta tacatatttt taaagaaaata ttataatgt atttatataa
 1081 tgtataaatg gttttatac caataaatgg catttaaaa aattc (SEQ ID NO: 11)

5

General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 62
- (2) 3' Untranslated Region - nts 699 - 1125

10

Initial Specific Target Motifs:

Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

6.5. Vascular Endothelial Growth Factor ("VEGF")

15

GenBank Accession # AF022375:

20

1 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcggcggt
 61 gcgagcagcg aaagcgacag gggcaaagtg agtgacctgc ttgtgggggt gaccggcgga
 121 gcgcggcggtg agccctcccc ctgggatcc cgcaagctgac cagtcgcgt gacggacaga
 181 cagacagaca ccgcggccag ccccaagttac cacccctcc cccgcggcg gcccacagtg
 241 gacgcggcggtg cgagccgcgg gcagggggccg gagccgcggcc cccggaggcg ggtggagggg
 301 gtcggagctc gccgcgtcgc actgaaactt ttctgtccaaac ttctgggtcg ttctcgcttc
 361 ggaggagccg tggccgcgc gggggaaagcc gagccgagcg gagccgcgag aagtgcgtac
 421 tcggccggg aggagccgca gcccggaggag gggggaggagg aagaagagaa ggaagaggag
 481 agggggccgc agtggcgact cggcgctcgg aagccggct catggacggg tgaggccggcg
 541 gtgtgcgcag acagtgcgtcc agcgcgcggc ctcccccagcc ctggccggc ctccggccgg
 601 gaggaaaggt agtcgcgcga ggccgcggagg agagccggcc gcccacagc ccgcggccgg
 661 gagggacgcg agccgcgcgc cccggcggtt cctccgaaac catgaactt ctgcgtctt
 721 ggggtgcattt gggccatttgc ttgtgcgtctt acctccacca tgccaaatgg tcccaaggctg
 781 caccatggc agaaggagga gggcagaatc atcacgaagt ggtgaagttc atggatgtct
 841 atcagcgcag ctactgcgtt ccaatcgaga cccctggatcatctccag gagtaccctg
 901 atgagatcga gtacatcttc aagccatctt gtgtgcgtt gatgcgtatgc gggggctgt
 961 ccaatgcgca gggccctggatgtgtgcgtt cttggggatgtc caacatcacc atgcagatgt
 1021 tgcggatcaa acctcaccaa ggcagcaca taggagatg gagctccata cagcacaaca
 1081 aatgtgaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggccctgt
 1141 cagagcggag aaagcatttgc ttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa
 1201 acacacactc gcggtgcgtt ggcaggcagc ttgagttaaa cgaacgtact tgcagatgt

30

35

35 General Target Regions:

- (1) 5' Untranslated Region - nts 1 - 701

(2) 3' Untranslated Region - nts 1275 - 3166

Initial Specific Target Motifs:

- 5 (1) Internal Ribosome Entry Site (IRES) in 5' untranslated region nts 513 -704
 5'CCGGGCUCAUGGACGGGUGAGGCAGCGUGUGCGCAGACAGU
 GCUCCAGCGCGCGCUCCCCAGCCCUGGCCGGCCUCGGGCCG
 GGAGGAAGAGGUAGCUCGCCAGGCCGGAGGGAGCAGGCCGCC
 CCCACAGCCCGAGCCGGAGAGGGACGCGAGGCCGCCGCCCCGU
 CGGGCCUCCGAAACCAUGAACUUUCUGCUGUCUUGGGUGCAUU
 GGAGCCUUGCUCUUGCUGCUACCACCAUG 3' (SEQ ID NO:
 13)
- 10 (2) Group III AU-Rich Element (ARE) Cluster in 3' untranslated region
 5' NAUUUAUUUAUUUAN 3' (SEQ ID NO: 10)

15

6.6. Human Immunodeficiency Virus I ("HIV-1")

GenBank Accession # NC_001802:

```

1 ggtctctctg gtagaccag atctgaggcct gggagcttc tggctaacta gggaaacccac
61 tgcttaagcc tcaataaaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt
121 gtgactctgg taactagaga tccctcagac ccttttagtc agtgtggaaa atctctagca
181 gtggcgcccg aacagggacc tgaaagcgaa agggaaacca gaggagctct ctgcacgcag
241 gactcggctt gctgaaggcgc gcacggcaag aggcgagggg cggcgactgg tgagtacgcc
301 aaaaaatttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa
361 gcgggggaga attagatcga tggaaaaaaa ttcggttaag gccaggggaa aagaaaaaat
421 ataaaattaaa acatatacgta tggcaagca gggagctaga acgattcgca gttatcctg
481 gcctgttaga aacatcgaaa ggctgttagac aaatactggg acagctacaa ccattccctc
541 agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc
601 atcaaaggat agagataaaa gacaccaagg aagctttaga caagatagag gaagagcaaa
661 acaaaagtaa gaaaaaagca cagcaaggcag cagctgacac aggacacagc aatcaggatca
721 gccaaaatta ccctatagtg cagaacatcc aggggcaat ggtacatcag gccatatcac
781 ctagaacttt aaatgcattgg gtaaaaggtag tagaagagaaa ggcttcagc ccagaagtga
841 taccatgtt ttcatcgatcca tcagaaggag ccacccaca agatttaaac accatgctaa
901 acacagtggg gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgagggaaag
961 ctgcagaatg ggatagagtg catccagtgc atgcaggccc tattgcacca ggccagatga
1021 gagaaccaag gggaaagtgc atagcaggaa ctactagtac ctttcaggaa caaataggat
1081 ggatgacaaa taatccaccc atccccatgg gagaatgg ataatccctgg

```

1141 gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac
1201 caaaggaacc ctttagagac tatgtagacc ggttctataa aactctaaga gccgagcaag
1261 cttcacagga ggtaaaaat tggatgacag aaaccttgtt ggtccaaaat gcgaacccag
5 1321 attgttaagac tattttaaaa gcatgggac cagcggctac actagaagaa atgtatgacag
1381 catgtcaggg attaggagga cccggccata aggcaagagt ttggctgaa gcaatgagcc
1441 aagtaacaaa tttagtacc ataatgtgc agagaggcaa tttaggaac caaagaaga
1501 ttgttaagtgt ttcaattgtt ggcaaaagaag ggcacacacg cagaaattgc agggccccta
1561 gaaaaaaggg ctgttgaaa tgtggaaagg aaggacacca aatgaaagat tgtactgaga
10 1621 gacaggctaa ttttttaggg aagatctggc cttcctacaa gggaggcca gggatttc
1681 ttcagagcag accagagccca acagccccac cagaagagag cttaggtct gggtagaga
1741 caacaactcc ccctcagaag caggagccga tagacaagga actgtatcct ttaacttccc
1801 tcaggtcact ctggcaac gaccctcgta cacaataaaat atagggggc aactaaagga
1861 agctctatta gatacaggag cagatgatac agtatttagaa gaaatgagtt tgccaggaag
1921 atggaaacca aaaatgatag gggaaattgg aggttttac aaagtaagac agtatgatca
1981 gatactcata gaaatctgtg gacataaagc tataggtaca gtatttagtag gac tacacc
2041 tgtcaacata attggaagaa atctgttgac tcagattggt tgcaatttaa atttcccat
2101 tagccattt gagactgtac cagtaaaattt aaagccagga atggatggcc caaaatttaa
2161 acaatggcca ttgacagaag aaaaaataaa agcatttagta gaaattttaa cagagatgga
2221 aaaggaaggg aaaatttcaa aaattgggcc tgaaaatcca tacaatactc cagttttgc
2281 cataaagaaa aaagacagta ctaaatggag aaaatttagta gatttcagag aacttaataa
2341 gagaactcaa gacttctggg aagttcaattt aggaataccat catccgcag gttaaaaaa
2401 gaaaaaatca gtaacagtac tggatgtggg ttagtgcataat tttcagttc ccttagatga
2461 agacttcagg aagtatactg catttaccat acctgtata aacaatgaga caccaggat
25 2521 tagatatcag tacaatgtgc tccccacaggg atggaaaggta tcaccagcaa tattccaaag
2581 tagcatgaca aaaatcttag agccttttag aaaacaaaaat ccagacatag ttatctatca
2641 atacatggat gatttgcatac taggtatctga ctttagaaata gggcagcata gaacaaaaat
2701 agaggagctg agacaacatc tggatgtggg gggacttacc acaccagaca aaaaacatca
2761 gaaagaacctt ccattccctt ggttgggtta tgaactccat cctgataat ggacagtaca
2821 gcctatagtgtc ctggccagaaaa aagacagctg gactgtcaat gacatacaga agttagtggg
2881 gaaattgaat tggcaagtc agatttaccc agggattaaa gtaaggcaat tatgtaaact
2941 ctttagagga accaaaggac taacagaagt aataccacta acagaagaag cagagctaga
3001 actggcagaa aacagagaga ttctaaaaga accagtacat ggatgttattt atgacccatc
3061 aaaagactta atagcagaaa tacagaagca gggcaaggc caatggacat atcaaattta
3121 tcaagagcca ttaaaaatc tgaaaacagg aaaatatgca agaatgaggg gtgcccacac
35 3181 taatgtatgtaa aaacaattaa cagaggcagt gaaaaata accacagaaa gcatagtaat

5341 gcagaatagg cgttactcgag cagaggagag caagaaatgg agccagttaga tcctagacta
5401 gagccctgga agcatccagg aagttagcctt aaaactgctt gtaccaattt ctattgtaaa
5461 aagtgttgct ttcatggca agtttgttc ataacaaaag ccttaggcatt ctcctatggc
5 5521 aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagctct
5581 ctatcaaagc agtaagtagt acatgtaatg caacctatac caatagtagc aatagtagca
5641 ttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaatatagg
5701 aaaatattaa gacaaagaaa aatagacagg ttaatgtata gactaataga aagagcagaa
5761 gacagtggca atgagagtga aggagaaata tcagcaccc tgtagatggg ggtggagatg
10 5821 gggcaccatg ctccctggga tggatgtatg ctgtgtgtc acagaaaaat tgggggtcac
5881 agtctattat ggggtacccg tgtggaaagga agcaaccacc actctatccc gtgcacatcaga
5941 tgctaaagca tatgatacag aggtacataa tggggcc acacatgcct gtgtacccac
6001 agaccccaac ccacaagaag tagtatttgtt aatgtgaca gaaaattttt acatgtggaa
6061 aaatgacatg gttagaacaga tgcacggata tataatcagt ttatggatc aaaggctaaa
15 6121 gccatgtgtaa atttaaccc cactctgtgt tagttaaag tgcactgatt tgaagaatga
6181 tactaatacc aatagtagta gcgggagaat gataatggag aaaggagaga taaaaaaactg
6241 ctcttcaat atcagcacaa gcataagagg taaggtgcag aaagaatatg catttttttta
6301 taaacttgat ataataccaa tagataatga tactaccagc tataagttaa caagttgtaa
6361 cacctcagtc attacacagg cctgtccaa ggtatccctt gagccattt ccatacattt
6421 ttgtcccccg gctgggtttcgattctaaa atgtataat aagacgtca atggAACAGG
6481 accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac
6541 tcaactgtcg ttaaatggca gtctagcaga agaagaggtt gtaatttagat ctgtcaattt
6601 cacggacaat gctaaaacca taatagtaca gctgaacaca tctgtgaaa ttaattgtac
6661 aagacccaaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagaggatt
6721 tggatcaataa gaaaaatag gaaatatgg acaagcacat tggatcaattt gtagagcaaa
6781 atggatataac actttaaaac agatagctag caaattaaga gaacaatttggaaataataa
6841 aacaataatc ttaaggcaat cctcaggagg ggacccagaa attgtaacgc acagttttaa
6901 ttgtggaggg gaattttctt actgttaattt aacacaactg ttaatgtttaa
6961 tagtacttgg agtactgaag ggtcaaaataa cactgaaggg agtgacacaa tcaccctccc
7021 atgcagaataa aacaaatataa taaacatgtg gcagaaagta ggaaaagcaa tggatcccc
7081 tcccatcagt ggacaaatataa gatgttcattt aaatattaca gggctgtat taacaagaga
7141 tggatcaataa gcaacaatg agtccggat cttcagaccc ggaggaggat atatggggaa
7201 caattggaga agtgaatttataa agtagtaaaa attgaaccat taggatgtac
7261 acccacaag gcaaaagagaa ggtggatc gagagaaaaa agagcgttgg gaataggagc
7321 ttgtccctt ggggttcgttgg gagcagcagg aagcactatg ggcgcagcctt caatgacgt
35 7381 gacggatcag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag

7441 ggctatttag ggcacaaacgc atctgttgc actcacagtc tggggcatca agcagctcca
 7501 gcaagaatc ctggctgtgg aaagataacct aaaggatcaa cagctccctgg ggaittgggg
 7561 ttgctctgga aaactcattt gcaccactgc tgtgccttgg aatgcttagtt ggagtaataaa
 5 7621 atctctggaa cagatttgga atcacacgac ctggatggag tggacagag aaattaacaa
 7681 ttacacaagc ttaatacact ccttaattga agaatcgca aaccagcaag aaaagaatga
 7741 acaagaattha ttggaatttag ataaaatgggc aagtttgtgg aattggttt acataacaaa
 7801 ttggctgtgg tatataaaat tattcataat gatagtagga ggcttggtag gtttaagaat
 7861 agttttgtgt gtactttcta tagtgaatag agttaggcag ggatattcac cattatcg
 10 7921 tcagacccac ctcccaaccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg
 7981 tggagagaga gacagagaca gatccattcg attagtgaac ggatccctgg cacttatctg
 8041 ggacgatctg cggagcctgt gccttcag ctaccaccgc ttgagagact tactcttgat
 8101 tgtaacgagg attgttggAAC ttctggac caggggggtgg gaagccctca aatattggtg
 8161 gaatctccta cagtatttggaa gtcaggaact aaagaatagt gctgttagct tgctcaatgc
 15 8221 cacagccata gcagtagctg aggggacaga tagggttata gaagtagtac aaggagcttg
 8281 tagagctatt cgccacatac ctagaagaat aagacaggc ttggaaagga ttttgtata
 8341 agatgggtgg caagtggtca aaaagttagt tgattggatg gcctactgt aggaaagaa
 8401 tgagacgagc tgagccagca gcagataggg tggagcagc atctcgagac ctggaaaaac
 8461 atggagcaat cacaagttagc aatacagcag ctaccaatgc tgcttgcc tggctagaag
 20 8521 cacaagagga ggaggagggtg gttttccag tcacacctca ggtacccctta agaccaatga
 8581 ctacaaggc agctgttagat cttagccact tttaaaaga aaaggggggc ctggaaggc
 8641 taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct
 8701 actccctga ttagcagaac tacacaccag ggccagggtt cagatatcca ctgacccttg
 8761 gatggtgcta caagcttagta ccagttgagc cagataagat agaaggaggc aataaaggag
 25 8821 agaacaccag ctgttacac cctgtgagcc tgcatggat ggatgaccgc gagagagaag
 8881 tggtagatg gaggtttgc agccgcctag catttcatac cgtggccgc gagctgcac
 8941 cggagttactt caagaactgc tgacatcgag ctgttacaa gggacttcc gctggggact
 9001 ttccagggag gctgtggcctg ggcgggactg gggagtggcg agccctcaga tccgtcatat
 9061 aagcagctgc ttttgccctg tactgggtct ctctggtag accagatctg agcctggag
 30 9121 ctctctggct aactaggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgc
 9181 c (SEQ ID NO: 14)

Initial Specific Target Motifs:

- (1) Trans-activation response region/Tat protein binding site - TAR RNA - nts 1
 35 - 60
 "Minimal" TAR RNA element

5' GGCAGAUCUGAGCCUGGGAGCUCUCUGCC 3' (SEQ ID NO: 15)

- (2) Gag/Pol Frameshifting Site - "Minimal" frameshifting element
5'

5 UUUUUUAGGGAAAGAUCUGGCCUUCUACAAGGGAAAGGCCAGG
GAAUUUUUCUU 3' (SEQ ID NO: 16)

6.7. Hepatitis C Virus (“HCV” - Genotypes 1a & 1b)

GenBank Accession # NC_001433:

1 ttggggcga cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga
61 aagcgtctag ccatggcggtt agtatgagtg ttgtgcagcc tccaggaccc cccctcccg
121 gagagccata gtggctcgcg gaaccggtga gtacaccgga attgccagga cgaccggcgt
181 ctttcctgga tcaacccgct caatgcctgg agatttgggc gtgccccgc gagactgcta
241 gccgagtagt gttgggtcgc gaaaggcctt gtggtaactgc ctgatagggt gcttgcgagt
301 gccccggag gtctcgtaga ccgtgcatca tgagcacaaa tcctaaacct caaagaaaaa
361 ccaaacgtaa caccaaccgc cgcccacagg acgttaagtt cccggcgggt ggtcagatcg
421 ttggtgaggat ttacctgtt ccgcgcaggg gccccaggtt gggtgtgcgc gcgacttagga
481 agacttccga gcggtcgcaa cctcgtaa ggcgacaacc tatcccaag gctcgccggc
541 ccgagggtag gacctggct cagccccgggtt acccttggcc cctctatggc aacgagggtt
601 tggggggc aggatggctc ctgtcacccc gtggctctcg gcctagtgg gcccccacag
661 acccccggcg taggtcgcgt aattttggta aggtcatcga tacccttaca tgcggcttcg
721 ccgacacctat ggggtacatt ccgcgttgcg ggcgcgcgcgcgcgcgcgcgcgcgcgc
781 tggcacatgg tgtccgggtt ctggaggacg ggcgtgaacta tgcaacagg aatctgccc
841 gttgccttt ctctatcttc ctcttagtt tgctgtcttg ttgaccatc ccagcttcg
901 cttacgaggt gcgcaacgtg tccggatata accatgtcac gaacgactgc tccaactcaa
961 gtatttgta tgaggcagcg gacatgtatca tgcacaccccg cgggtgcgtg ccctgcgtcc
1021 gggagagtaa ttctccgt tgctggtagt cgctactcc cacgctcgcg gccaggaaca
1081 gcagcatccc caccacgaca atacgacgcc acgtcgattt gtcgttggg gcccgtgc
1141 tctgtccgc tatgtacgtt ggggatctcg cggatccgtt tttctgtc tcccagctgt
1201 tcacccctc acctcgccgg tatgagacgg tacaaggatgg caattgtctca atctatcccg
1261 gcccgtatc aggtcaccgc atggcttggg atatgtatgg gaaactggtca cctacaacgg
1321 cccttagtggt atcgcagcta ctccggatcc cacaaggccgt cgtggacatg tggccgggggg
1381 cccactgggg tgccttagcg ggcctgcct actattccat ggtggggaaac tgggctaagg
1441 tcttgattgt gatgtactc ttgtctggcg ttgacggca caccacgtg acagggggaa
1501 gggtagcctc cagcacccag agcctcggt cctggcttc acaaggccca tctcagaaaa
1561 tccaaactcgtaa acacccacac ggcagctggc acatcaacag gaccgcgtcg aatttgcataat

1621 actccctcca aactgggttc attgctgcgc tggtaacgc acacagggtc aacgcgtccg
 1681 ggtccccaga gcgcattggct agctgccgc ccatcgatga gttcgctag gggtggggtc
 1741 ccatcaactca tgatatgcct gagagctcg accagaggcc atattgctgg cactacgcgc
 5 1801 ctgcaccgtg cgggatcgta cttgcgtgc aggtgtgtgg tccagtgtat tgcttcactc
 1861 cgagccctgt ttagtgtggg acgaccgatc gttcggcgc tcctacgtat agctgggggg
 1921 agaatgagac agacgtgctg ctacttagca acacgcggcc gcctaaggc aactggtttg
 1981 ggtgcacgtg gatgaacagg actgggtca ccaagacgtg cggggggccct ccgtgcaaca
 -2041 tcgggggggt cggcaacaac accttggctc gccccacgga ttgctccgg aagcaccccg
 10 2101 aggccactta cacaaagtgt ggctggggc cctgggtjac acccaggtgc atgggtgact
 2161 acccatacag gctctggcac taccctgca ctgttaactt taccgtctt aaggtcagga
 2221 ttagtgtggg gggcgtggag cacaggctca atgctgcatg caattggact cgaggagagc
 2281 gctgtgactt ggaggacagg gataggtcag aactcagccc gctgctgctg tctacaacag
 2341 agtggcagat actgcctgt tccttacca ccctaccggc cctgtccact ggcttgatcc
 15 2401 atcttcacccg gaacatcgta gacgtgcaat acctgtacgg tatagggtcg gcagttgtct
 2461 ctttgcata caaatgggag tataatgcctgt tgcctttcct tcttctggcg gacgcgcgcg
 2521 tctgtgcctg ctgtggatg atgctgcata tagcccaggc tgaggccacc ttagagaacc
 2581 tggtgtcct caatgcggcg tctgtggccg gagcgcataa ctttctctt ttctctgt
 2641 tcttctgcgc cgcctggatc atcaaaggca ggctggccc tggggccgc tatgtctct
 2701 atggcgtatg gccgttgctc ctgtcttgc tggccattacc accacgagct tatgccatgg
 2761 accgagagat ggctgcatacg tgcggaggcg cggttttgt aggtctggta ctcttgac
 2821 tgcaccata ctataagggt ttccctgcata ggctcatatg gtggttacaa tattttatca
 2881 ccagagccga ggcgcacttg caagtgtggg tccccctct caatgtcgg ggaggccgcg
 2941 atgccatcat cttccatata tgcgcggtcc atccagagct aatcttgac atcaccaa
 3001 tcctgtcgc catactcggt ccgcctatgg tgctccaggc tggcataact agagtgcgcgt
 3061 actttgtacg cgctcagggg ctcatccgtg catgcattt agtgcggaaag gtcgtggag
 3121 gcaactatgt ccaaattggcc ttcatgaagc tggccgcgt gacaggtagc tacgtatatg
 3181 accatcttac tccactgcgg gattggggcc acgcgggcct acgagaccit ggcgtggcag
 3241 tagagcccgatgt cgtttctct gacatggaga cttaactcat cacctggggg gcagacaccg
 3301 cggcgtgtgg ggacatcatc tcgggtctac cagtcgcgc ccgaaggggg aaggagatac
 3361 ttcttaggacc ggccgatagt ttggagagc aggggtggcg gtccttgcg cctatcacgg
 3421 cttatccca ccaaacgcgg ggcctgcctt gctgtatcat cactagccct acaggtcggg
 3481 acaagaacca ggtcgatggg gaggttcagg tgcctccac cgcaacgcaa tcttcctgg
 3541 cgacactgcgt caatggcgta tggtggaccg tctaccatgg tgccggctcg aagaccctgg
 3601 cggcccgaa gggccaatc acccaaatgt acaccaatgt agaccaggac ctgcgtggct
 3661 ggcggccgc ccccgggccg cgtccatga caccgtgcac ctgcggcagc tcggaccc

3721 acttggtcac gaggcatgct gatgtcgltc cggtgcgcgc gcggggcgac agcaggggga
3781 gcctgcttc ccccaggccc atctcctacc tgaagggctc ctgggtgga ccactgcttt
3841 gccctcggg gcacgttcta ggcattcc gggctgtgt gtgcacccgg ggggttgcga
5 3901 aggccgtgga cttcataccctt gttgagtcta tggaaactac catgcggctt ccggcttca
3961 cagacaactc atccccctcg gccgttaccgc aaacattcca agtggcacat ttacacgctc
4021 ccactggcag cggcaagagc accaaagtgc cggctgcata tgcaagccaa gggtaacaagg
4081 tgctcgctt aaacccgtcc gttgccgcca cattgggtt tggagcgat atgtccaagg
4141 cacatggcat cgagcctaaccatcagaactg gggtaaggac catcaccacg ggccggccca
10 4201 tcacgtactc cacctattgc aagttccttg ccgacgggtt atgtccggg ggccctatg
4261 acatcataat atgtgatgaa tgccactcaa ctgactcgac taccatctt ggcatacgca
4321 cagtcctgga tcaggcagag acggctggag cggcgtcgat cgtgcgtcc accggcacgc
4381 ctccggatc gatcaccgtg ccacacccca acatcgagga agtggccctg tccaacactg
4441 gagagattcc cttctatggc aaagccatcc ccattgaggc catcaagggg ggaaggcatc
15 4501 tcatacttctg ccattccaag aagaagtgtg acgagctcg cggcaagctg acaggcctcg
4561 gactcaatgc ttagcgat taccggggtc tcgatgtgc cgtcataccg actagcggag
4621 acgtcggtt cttggcaaca gacgctctaa tgacgggtt taccggcgac ttgactcag
4681 ttagtcgactg caacacatgt gtcacccaga cagtcgattt cagctggat cccaccctca
4741 ccattgagac gacaacgctg ccccaagacg cgggtgcgc tgccagcgg cgaggttaga
20 4801 ctggcagggg caggagtgcc atctacaggt ttgtgactcc aggagaacgg ccctcaggca
4861 ttttcgactc ctcggctctg ttgtgactgt atgacgcagg ctgcgttgg tatgagetca
4921 ccccccgtga gacctcggtt aggtgcggg cttacctaaa tacaccaggg ttggccgtct
4981 gcccaggacca ccttaggttc tgggagagcg tttcacagg cctcacccac atagatgccc
5041 acttcttgc ccagacaaa caggcaggag acaaccccctt acactggta gcataccaag
25 5101 ccacagtgtg cggccagggtt caggctccac ctccatgtt ggaccaaatg tggaaagtgtc
5161 tcatacggct aaagccccaca ctgcattggc caacgcccctt gctgtacagg ctaggagccg
5221 ttcaaaatga ggtcaactctc acacacccca taaccaataa catcatggca tgcatgtcgg
5281 ctgacccgtt ggtcgactt acgacccgtt tgcttagtgg cggagtcctt gcccgtctgg
5341 cccgtactg cctgacgaca ggcagcgtgg tcatttggg caggatcatc ttgtccggga
30 5401 gcccaggctgt tattcccgac agggaaagtcc tctaccagga gttcgatgag atggaaagagt
5461 gtgttcaca ctccttac atcgagcaag gaatgcagct cggcggccaa ttcaaacaga
5521 aggcgcgtgg attgctgcaaa acagccacca agcaagcgga ggctgtcgct cccgtggfgg
5581 agtccaaatgt gcgagccctt gaggtcttctt gggcgaaaca catgtgaaac ttcatcagcg
5641 ggatacagta cttggcaggc ctatccactc tgccctggaaa ccccgccata gcatcattga
35 5701 tggctttac agccttatac accagccgc tcaccacccaa aataaccctc ctgttttaaca
5761 tcttgggggg atgggtggct gcccacgtt ctcggccag cgtgttgcgtt gctttcggtt

5821 gcccggcat tgccggcg gccgtggca gcataggctc cggaaaggta cttgtggaca
5881 ttcggcggg ctatgggcg ggggtggctg ggcactcgt ggccttaag gtcgtggcg
5941 gcgagatgcc ctccactgag gatctggta atttactccc tgccatccct tctctggcg
5 6001 ccctggtgtt cggggtcgtg tgccgaccaa tactgcgtcg gcacgtggc cggggagagg
6061 gggctgtgca gtggatgaac cggctgatag cggtcgctc gcggggtaac cacgtctccc
6121 ccacgcacta tgtgcccggag agcgacgccc cggcgcgtgt tactcagatc ctctccagcc
6181 ttaccatcac tcagtgcgt aagaggctc atcagtggat taatgaggac tgctccacgc
6241 cttgtccgg ctctgtggcta aaggatgtt gggactggat atgcacgtgt ttgagtgact
10 6301 tcaagacttg gtcctcactc aagctctgc cgccgttacc gggactccct ttctgtcat
6361 gccaacgcgg gtacaaggga gtcggcggg gggatggcat catgcaaacc acctgccccat
6421 gtggagcaca gatcaccggaa catgtcaaaa atggctccat gaggattgtt gggccaaaaa
6481 cctgcagcaa cacgtggcat ggaacattcc ccatcaacgc atacaccacg gcccctgca
6541 cgcctcccc agcgccgaac tattccaggcg cgctgtggcg ggtggctgt gaggagtacg
15 6601 tggaggttac gcgggtggg gatttccact acgtgacggg catgaccact gacaacgtga
6661 aatgcccattc ccagggttcca gcccctgaat ttacggat ggtggatggaa gtacgggtgc
6721 acaggatgc tccagtgatc aaacccctcc tacgagagga ggtcgatattc caggcggc
6781 tcaaccagta cctggcggg tcacagctcc catgtgagcc cgaaccggat gtggcagtgc
6841 tcactccat gtcaccggac ccctctata ttacagcaga gacggccaag cgtaggctgg
20 6901 ccagggggtc tccccctcc ttggccactt ccagggttgc ggccttcc
6961 tgaaggcgac atgtactacc catcatgact cccggacgc tgacccatc gaggccaaacc
7021 tcctgtggcg gcaggagat ggccggaaaca tcacccgtgt ggagttagaa aataagggtgg
7081 taatcccttgcgat ccgttccggg cgggtggagga tgagaggaa atatccgtcc
7141 cggcggagat cctgcgaaaa cccaggaatg tccccccagc gttggccata tgggcacgcc
25 7201 cggattacaa ccctccactg cttaggttcc ggaaggaccc ggactacgtc ccccccgggtgg
7261 tacacgggtg cccttgcgc tctaccaagg ccccccata accaccatcc cggagggaa
7321 ggacgggtgt cctgacagag tccaccgtgt ttctgcctt ggcggagctc gctactaaga
7381 ctttggcag ctccgggtcg tcggccgtt acagcggcac ggcgactggc cctcccgatc
7441 aggccctccga cgacggcgac aaaggatccg acgttggatc gtactccatc atgccccccc
30 7501 tcgagggaga gccaggggac cccgacctca ggcacgggtc ttggctacc gtgagcgggg
7561 aagctggatgaa ggacgtcgat tgctgttccaa tgccttatac atggacaggt gcttgcgtca
7621 cggcatgcgc tgccggaggag agcaagtgc ccatcaatcc gttggacaaac tcttgcgtgc
7681 gtcaccacag tatggtctac tccacaacat ctcgcagcgc aagtctgcgg cagaagaagg
7741 tcacccatgaa cagactgcaatg tcctggacg accactaccg ggacgtgtc aaggagatga
35 7801 aggcgaaggc gtccacagtt aaggcttaggc ttctatctat agaggaggcc tgcaaaactga
7861 cggcccccaca ttggccaaa tccaaatting gtcacggggc gaaggacgtc cggagccat

7921 ccagcaggc cgtcaaccac atccgctccg tgtggagga ctgctggaa gacactgaaa
 7981 caccaattga taccaccatc atggaaaaaa atgaggttt ctgcgtccaa ccagagaaag
 8041 gaggccgcaa gccagctcg cttatcgat tcccagacct ggggtacgt gtatcgaga
 5 8101 agatggccctt tacgacgtg gtctccaccc ttccctcaggc cgtgatggc ccctcatacg
 8161 gattccagta ctctcctggg cagcgggtcg agttcctggt gaatacctgg aaatcaaaga
 8221 aatgccctat gggcttctca tatgacaccc gctgcttga ctcaacggc actgagaatg
 8281 acatccgtac tgaggaatca atttaccaat gtttgactt ggccccgaa gccaggcagg
 8341 ccataaggc gtcacagag cggctttag tcgggggtcc cctgactaat tcgaagggc
 8401 agaactgcgg ttatcgccgg tgccgceaa gtttgtcgact gacgactagc tgccgcaaca
 10 8461 ccctcacatg ttacttgaag gccactgcgg cctgtcgac tgcaaaagtc caggactgca
 8521 cgtatcgctgtt gAACGGGAGAC GACCTTGTGTT TATCTGTGA GAGTGCGGGAA ACCCAGGAGG
 8581 atgcggcggc cctacgagcc ttacggagg ctatgactag gtattccgccc cccccccgggg
 8641 acccgccccca accagaatac gacttggagc tgataacgac atgctctcc aatgtgtcg
 15 8701 tcgcgcacga tgcatccggc aaaagggtt actacccatc ccgtggcccc acccccccc
 8761 tcgcacgggc tgctgggag acatggtagac acactccagt caactctgg ctaggcaata
 8821 tcatcatgtatgcggccacc ctatggcga ggtatgattt gatgactcat ttctctcta
 8881 tcctcttagc tcaggagcaa cttgaaaaag ccctggattt tcaataggtt gggcctgtt
 8941 actccattga gccacttgcac ctacccatc tcaatttgcg actccatgtt cttggccat
 9001 ttcaactcca cagttactt ccaggtgaga tcaataggtt ggcttcatgc ctcaggaaac
 20 9061 ttggggtacc gccttgcga gtctggagac atcggggccag aagtgtccgc gctaaagctac
 9121 tggcccgagg ggggagggtt gcccacttgcg gcaaggatctt cttcaactgg gcagtaaaga
 9181 ccaagcttaa actcaactcca atcccgctg cgtcccgact agacttgcctt ggctgggtcg
 9241 ttgtggta caacggggga gacatatatc acaggctgc tctgtccggc ccccggttgg
 9301 tcatgttgc cttacttca ctttctgtt gggtaggcac tttttttttt ttc (SEQ ID NO: 17)
 25 9361 gaacggggag ctaaccactc caggccaata ggccattccc tttttttttt ttc (SEQ ID NO: 17)

General Target Region:

5' Untranslated Region - nts 1 - 328 - Internal Ribosome Entry Site (IRES):

30 5'UUGGGGGCGACACUCCACCAUAGAUUCACUCCCCUGUGAGGAACUACUGUCU
 UCACGCAGAAAGCGUCUAGCCAUGGCGUUAGUAUGAGUGUUGUGCAGCCUC
 CAGGACCCCCCCCUCCCGGGAGAGCCAUCAGUGGUUCGUUGGAUCAACCCGCUAAUGC
 ACCGGAAUUGCCAGGACGACCGGGGUCCUUUCUUGGAUCAACCCGCUAAUGC
 CUUGGAGAUUUGGGCGUGGCCCGCGAGACUGCUAGCCGAGUAGUGUUGGGU
 CGCGAAAGGCCUUGGGUACUGCCUGAUAGGGUGCUUGCGAGUGUGCCCCGGG
 35 AGGUCUCGUAGACCGUGCAU3' (SEQ ID NO: 18)

Initial Specific Target Motifs:

- 5 (1) Subdomain IIIc within HCV IRES - nts 213 - 226
 5'AUUUGGGCGUGCCC3' (SEQ ID NO: 19)
- 10 (2) Subdomain IIId within HCV IRES - nts 241-267
 5'GCCGAGUAGUGUUGGGUCGCGAAAGGC3' (SEQ ID NO: 20)

6.8. Ribonuclease P RNA ("RNaseP")

10 GenBank Accession #s

X15624 Homo sapiens RNaseP H1 RNA:
 1 atggcgagg ggaagctcat cagtggggcc acgagctgag tgcgtctgt cactccactc
 61 ccatgtccct tggaaaggtc tgagactagg gccagaggcg gccctaacag ggctctccct
 121 gagcttcagg gaggttagtt cccagagaac ggggctccgc gcgaggctag actggcagg
 181 agatgccgtg gacccccc ttccccgggg ggcccgccgg atgcctccct tgccggagct
 241 tggAACAGAC tcacggccag cgaagttagt tcaatggctg aggtgaggta ccccgccagg
 301 gacctataaa cccaatttag accactctcc tccgcccatt (SEQ ID NO: 21)

U64885 Staphylococcus aureus RNaseP (rrnB) RNA:
 1 gaggaaaagtc cgggctcaca cagtctgaga tgattgttagt gtcgtgctt gatgaaacaa
 61 taaatcaagg cattaatttg acggcaatga aatatcctaa gtcttcgtat atggatagag
 121 taatttgaaa gtgccacagt gacgttagtt ttatagaaat ataaaagggtg gaacgcggta
 181 aacccctcga gtgagaatac caaattttgtt aggagcacctt gtttaacgga attcaacgta
 241 taaacgagac acacttcgca aatgaagtg gtgttagacag atggttatca cctgagtacc
 301 agtgtgacta gtgcacgtga tgagtacgt ggaacagaac gcggcttat (SEQ ID NO: 22)

M17569 Escherichia coli RNA component (M1 RNA) of ribonuclease P
 (rnpB) gene:
 1 gaagctgacc agacagtgcg cgcttcgtcg tgcgtctttt cgggggagac gggcgagg
 61 gaggaaaagtc cgggctccat agggcagggt gccaggtaac gcctgggggg gaaacccacg
 121 accagtgc当地 cagagagcaa accggccatg gcccgc当地 gcgggatcag gtaagggtga
 181 aagggtgc当地 taagagc当地 ccgc当地ggct ggtaacagtc cgtggcacgg taaactccac
 241 ccggagcaag gccaatagg ggttcataag gtacggcccg tactgaaccc ggttaggctg
 301 cttagccag tgagcgattt ctggcttaga tgaatgactg tccacgacag aaccggctt
 361 atcggtc当地 ttccac (SEQ ID NO: 23)

Z70692 Mycobacterium tuberculosis RNaseP (rnpB) RNA:

1 ccacccgtta cgatcttgcc gaccatggcc ccacaatagg gccggggaga cccggcgta
 61 gtggtggcgc gcacggtcag taacgtctgc gcaacacggg gttgactgac gggcaata
 5 121 ggctccatag cgtcgccgc ggatacaga aaggagcatt ctgtgacgga aaagacgccc
 181 gacgacgtct tcaaacttgc caaggacgag aaggtcgaat atgtcgtacgt ccgggtctgt
 241 gacctgcctg gcatcatgca gcacccacg attccggctt cggccttga caagagcgtg
 301 ttggacgacg gcttggcctt tgacggctcg tcgattcgcg ggttccagtc gatccacgaa
 361 tccgacatgt tgcttctcc cgatcccgag acggcgccca tcgaccgtt ccgcggcc
 421 aagacgctga atatcaactt ctttgtcac gaccgttca ccctggagcc gtactccc
 10 481 gaccgcgca acatcgcccg caaggccgag aactacctga tcagcactgg catcgccgac
 541 accgcatact tcggcgccga ggccgagttc tacatttgcg attcggtgag cttcgactcg
 601 cgcgccaacg gcttccatca cgaggtggac gccatctcg ggtggggaa caccggcg
 661 gcgaccgagg ccgacggcag tcccaaccgg ggctacaagg tccgccacaa gggcggttat
 15 721 ttcccagtgg cccccaacga ccaatacgtc gacccgtcg acaagatgct gaccaacctg
 781 atcaactccg gcttcatctt ggagaaggcc caccacgagg tggcagcgg cggacaggcc
 841 gagatcaact accagttcaa ttgcgtctg cacggcccg acgacatgca gtttacaag
 901 tacatcatca agaacaccgc ctggcagaac ggcaaaacgg tcacgttcat gccaagccg
 961 ctgtcggcg acaacgggtc cggcatgcac tgtcatcagt cgctgtggaa ggacggggcc
 1021 ccgctgatgt acgacgagac gggttatgcc ggttgcgtcg acacggcccg tcattacatc
 1081 ggcggcctgt tacaccacgc gccgtcgctg ctggccttca ccaacccgac ggtgaactcc
 1141 tacaagccgc tgggtcccg ttacgaggcc cggatcaacc tggctatag ccagcgcaac
 1201 cggtcggcat gcgtgcgc cat cccgatcacc ggcagcaacc cgaaggccaa gggctggag
 1261 ttccgaagcc ccgactcgac gggcaacccg tatctggctg tctcggccat gctgatggca
 1321 ggcctggacg gtatcaagaa caagatcgag ccgcaggcgc ccgtcgacaa ggatcttac
 1381 gagctccgc cgaaagaggc cgcgagtatc ccgcagactc cgacccagct gtcagatgt
 1441 atcgaccgtc tcgaggccga ccacgaatac ctcacccgaag gagggtgtt cacaacgac
 1501 ctgatcgaga cgtggatcgt ttcaagccgc gaaaacgaga tcgagccgtt caacatccgg
 1561 ccgcattccct acgaattcgc gctgtactac gacgttaag gactcttcgc agtccgggt
 1621 tagagggagc ggcgtgtcg tgcaggcgc ggcgtcgagg ttttcgtat ggtgacgggt
 1681 gccggcaacg ggcgcgcac caccgtcgac aagagccgt ttaagaacgt tcaaggacgt
 1741 ttccggccgg tgcaccaacc cgcgtggccaa tcatctcccg accgcgcgac ggggtgtt
 1801 tcacatgcgc cggaaactcaa gccacgtcg cgcggccgcg tgcgtcgac gccgggtc
 1861 gttaaatgtc ggggattcgt cgtgcggccgc ggcgtccacg ctgaccaacg gggcgtca
 1921 ctcccgaaaca ctttgcgcac taccgcctt gcccggccgc tcacccgttag gtagttgtcc
 35 1981 aggaattcccc caccgtcgac gtttgcgcag cggccgcga ccgcgaccgc attgagctgg

2041 cgcccgggtc cggcagctg gtcggggc ttgccgcga ccaacaccag cgcggtcg
2101 gcccggtgg cggcagcca ggcctgacgg agcagctca cgtcggtgc gggaaaccaga
2161 tcggcggccg cgatgacatc caggattgc agcgtcgagg tgggtgcag ggccggaaacc
2221 tgggtgcgtat gctgttagctg cagcaactgc acggtccatt cgatgtcgcc cagtccggccg
5 2281 cggcccgat tgggtgtgt gtggggcgc gcaccgcgc gcaaccgctc ggactcgata
2341 cggcccttga tgccgcgaat ctgcgcacc gagtcagcgg acacaccgtc gggcggatac
2401 cgcgtttgt cgaccatccg taggaatcg tgacccaact cggcatcgcc ggcaaccgcg
2461 tgtgcgcgtat gcagggcctg gatctccat ggctgtgccc actgctcgta gtatgcggcg
10 2521 taggacccca ggggtgcggac cagcggaccg ttggggccct cgggtcgcaa attgggtcg
2581 agctcccgcg cggatcgac gctgggtgc cccagcagcg cccgaaccccg ctgcggcgtatc
2641 gatgtcgacc attcaccgc cctgcgtatcg tcgacgcgg tggccggctc acagacgaac
2701 atcacgtcg catccgaccc gtagccaaac tcggcaccac ccagccgacc catgccgatg
2761 accgcgtatgg ccggccggggc gcgatcgac tcgggaaaggc tggccggat catgacgtcc
15 2821 agcgcggcct gcagcaccgc caccacacc gacgtcaacg cccggcacac ctgcgtgacc
2881 tcgagcaggc cgaggcggc cggccaaaccg atgcgggcca gctctcgacg acgcagcgatg
2941 cgcgcggccg cgatggcccg ctccgggtcg gggtagcggc tcggcggagc gatcagcggcc
3001 cgagccacgg cggcgggctc ggttcgagc agttcggc cccgaggccc gtcctcgatc
3061 tgctggatga cccgcggcgc ggcgtatcaac agatccggca catacgccga ggtacccaaag
3121 acatgcgtatga gcccgtggc caccgcggc ttgtccgcga gcggtggccag gtaccagctt
20 3181 tcggtgccca ggcctcaact gagccgcgg taggcccggca gtccgcgtc gggatcgaaaa
3241 gcatacgaca tccagtcag cagcctggc agcagcaccg actgcaccccg tccgcggccgg
3301 ccgcgttgcgtat tgaccaacgc cgacatgtgt ttcaacgcgg tctgcggatcc ctcgtagccc
3361 agcgcggcca gcccggccc cgcggctcc aacgtcatgc cgtggcgat ctccaaacccg
3421 gtcggccga tcgattccag cagcgggtga tagaagagtt tgggtgttaa ctgcacacc
25 3481 cgcacgttct gcttcgtatgc ttccctccgc agcaccggc cgcgtatcgat tcggccatcg
3541 gggccggatgt gggccgcgc cgcggccag cgcactgcct cctcgatc gggatcgaaa
3601 agcagggtggg tgcgttgcgtatgc cgcgtcaac tgcgtcggt gctcgagcag cctgaggaac
3661 tcatacgacg cggcatgtt cgcgcgtcc tcacgcggc tggtagccgc ttccggccaaac
3721 gcccggccatg cgtccaccgt ggacgcccacc cgtaaacgact cgtcgatcg ggcgtatcg
30 3781 agctgcgtatgc gctgtacggc gaactccacg tcgcgtatgc cgcgtcgcc gagtttgcgt
3841 tcgcggccgc ggacatcgcc gggcaccacg tgctccaccc gcccgcgtat gcgcgtcc
3901 tcgaccacaa agtccgcgtatgc ctcgcaggct cgcgtatcgaa tcggcatcaa ggcgggtcagg
3961 taacgcgtatgc caagttccgc gtcgcgtatgc actggccgtatgc tttcgtatgc cgcgtaaac
4021 tcccgaggatgt tggcccgatgc ctggtagtag gcgatgtcgactcgtatgc acggaccagc
35 4081 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4141 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4201 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4261 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4321 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4381 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4441 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4501 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4561 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4621 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4681 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4741 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4801 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4861 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4921 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
4981 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5041 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5101 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5161 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5221 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5281 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5341 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5401 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5461 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5521 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5581 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5641 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5701 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5761 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5821 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5881 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
5941 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6001 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6061 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6121 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6181 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6241 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6301 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6361 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6421 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6481 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6541 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6601 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6661 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6721 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6781 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6841 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6901 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
6961 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7021 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7081 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7141 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7201 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7261 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7321 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7381 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7441 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7501 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7561 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7621 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7681 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7741 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7801 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7861 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7921 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
7981 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8041 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8101 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8161 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8221 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8281 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8341 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8401 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8461 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8521 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8581 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8641 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8701 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8761 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8821 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8881 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
8941 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9001 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9061 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9121 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9181 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9241 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9301 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9361 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9421 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9481 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9541 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9601 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9661 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9721 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9781 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9841 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9901 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg
9961 tcccggtgc gcccctccgg acgcaggcggc ggcgtccaccc cggatcgatgc cgcgtatcg

4141 acccgcatca tctcgctggc cacgcgcgac ttgcgcgggt cggagcgctc ggcaacgaat
4201 atgacatcaga cgtcgctgac gtagttcagt tcgcgcac cgcaactgcc catcgatg
4261 accggccaggc gcggtggcggt gtgcgcggc cacacgctcg cctggccac gcgcagcgcc
5 4321 gcccgcagag cggcgccggc ggcgtccgcc aggcggtgcgg ccaccacggt gaatggcagc
4381 accgggtcgt cctcgaccgt cgccggcagg tcgagagcgg ccagcattag cacgtatcg
4441 cgttactggg ttcgcaatcg gtgcacgagc gagccggca taccctccga ttccctcgacg
4501 cactcgacga acgaccgcgt cagctggta tgggacggca gtgtgacattt gccccgcagc
4561 aatttccagg actgcggatg ggcgaccagg tgatcgccca acgcccagcga cgagcccagc
10 4621 accgagaaca gcccggcggc cagactcggt tcgcgcagca gagccgcgtt gagctcgcc
4681 catccggtgt ctggattctc cgacagccgg atcaaggcgc gcagcgcggc atccggcgtcc
4741 ggagcgcgtg acagcgacca cagcaggtcg acgtgcgcct gatccctcgat ccgatcccac
4801 cccagctgag ccagacgctc accagcaggg gggtaacta atccgagccg gccaacgcgt
4861 gcaacttcg gccgctgcgt ggcgagttt gtcacgacca cgacggtagc gcaaaggcgc
15 4921 tcggcgtcggt atcaaccggt agatctggc tacagcgaca ggttagtgcg cagctcgat
4981 ggctgacgt ggctgcggta gttcgccac tccgtgcgt tttgcgcaaa gaaaaagtca
5041 aaaacgtgtc ccccaaggc ctccgcacg agtgcggagg cttccatggc ggcagcgc
5101 ctatccaaac tggacggcaa ttctcggtac cccatcgctc ggcttcctc ggggtgagg
5161 tcccatatcg tgcctcggtc ctgcgggccc agcacgtaa cttctctac acccccaat
20 5221 cccggggcca gcagcacggc gaatgtcaga tagggattgc acggcaatc agggctgcgt
5281 acttcgaccc gccgcgcacg ggtttgtgc ggcgtgtaca tcggcacccg cactagggcg
5341 gatcggttgg cggcccccac caacgcgcgtt gttggcgctt cggccgcgtt caccagccgc
5401 ttgttaagagt tgacccactg atttgtgacc ggcgtgtat cgcaagcggtc ctccaggatc
5461 cggcgatgtc acgatttacc cacttcgcac agctgcagcg gatcatcagc gctgtggaaac
25 5521 gcgttgcacat caccctcgaa caggctcatg tgggtgtcga tcgcgcagcc cgggtgcgt
5581 ccgaatggct tgggcataa caacgcggc ggcgcctt ccagcgcac ttcttgcgt
5641 acgtacggc aggtcatcac ttgtcagcc atcgacagag cgtcgccaaa cggcaggatcg
5701 atctcctgcgtc ggccgggtgc gccttcgttca tggctgaact ccaccgagat gcccataat
5761 tccaggccat cgatcggttgc gcccggaaat ttcaaggcgg agtcgtgcac cgcttggcgt
30 5821 aaatagccgg cgttgcac cgggacgggc accgaccgt ctcgggtcc gggcttgagg
5881 aggaagaact cgatttcggg atgcacgtat caggagaagc cgatcgcc ggccttcgtc
5941 agctgcgcgc gcaacacgtt cccgggttcc gcccacgcac ggcgcgttcc cggcatggtg
6001 atgtcgcaaa acatccgcgc tgagtgggg tggccggaaac tggggccca gggcagcacc
6061 tggaaaggatcg acgggtccgg gtgcgcaccat gtatcggtt ccgagaccccg cgcaaaaggccc
35 6121 tcgatcgagg atccgtcgaa ggcgtgcct tcctcgagg cgccctcgag ttccgggtgg
6181 gcgatggcga ccgacttgat gaaaccgagc acgtcgatcgacc accacagccg gacgaaggcgg

6241 atgtcgctt cttccagggt acgaagaacg aattcctct gtccgtccat acctcgaaaca
6301 gtatgcactg tctgtaaaaa ccgtgttacc gatgccggc cagaagcggt gcggggcgcc
6361 ccgcaagggg agtgcgcggg gagttcaggg cgccgcaccgc agactcgctcg gcggcaaggt
5 6421 cccgtcgaga aaatagtgc a tcaccgcaga gtccacacac tggttccat cgaacaccgc
6481 agtgtgttgg gtgccgtcga aggtgatcag cggtgccccc agctggcggt ccaggtctac
6541 cccggactga tacggagtgg ccgggtcggt ggtggtggac accacgacga cttggccagc
6601 cccggccggc gccgcggggt gcggcgtcga cgtggccggc accggccaca gcgcgcacag
6661 atcgccccggg gcggatccgg tgaactgccc gtagctaagg aacggggcga cctgacggat
10 6721 ccgttggtcg gcggccaccc aggccgtgg atcgccgggt gtggcgcat cgacgcaccc
6781 gaccgcgttg aacgcgtctt ggtcggtgt gtagtgcggc tctgcattccc ggccgtcata
6841 gtgcgtggca agcaccagca agtcgcggc gtgcgtggc cgctgcagcc ccagcagacc
6901 actggtcagg tactccagc gctgagggt gtacagcgcg ttgatggtgc ccgtcgctgc
6961 gtccgcgttag ctccaggccac gtggatccga cgtcttaccc ggcttgcga ccagcgggtc
15 7021 aaccaggggcg tggtagcggt tgacccactg ggccgagtcg gtgcggagag ggccaggccgg
7081 cgagcggcg cagtcggcg cgtagtcatt gaaagggtc taaaatcccg ccattggct
7141 gatgtttcc tcgattgggc taacggctgg atcgatagcg ccgtcgagga ccatcgcccc
7201 cacatgagta ccgaaccgtt ccaggtaaac ggtgcggcaac tcggtgccgt agctgtatcc
7261 gaggtatgtt atctgtatcg cacctaacgc ttggcgaacc atgtccatgt cccgtgcgac
20 7321 ggacgcggta ccgatattgg ccaagaagct gaagcccatc cggtaaacac agtctggc
7381 caactgcgg tagacctgtt cgacgtgggt gacacggcc ggactgttagt cggccatcgg
7441 atcgcgcggg tacgcgtcga actcggcgtc ggtgcgacac cgcaacgcag gggtcgagtg
7501 gccgaccctt ctccgggtcgaa agcccaccag gtgcgaagtgg cggagaatgt cgggtcgcc
7561 gatcgccggg gccatagcgg cgaccatgtc gaccgcgcac gccccgggtc cccaggatt
25 7621 gaccagcgt gctccgaatc gctgtccgtt cggggggacg cggatcaccgc ccaacttcgc
7681 ttgttccca ccgggttgggt cgtagtcgac ggggacggac accgtcgccgc agctgtcgat
7741 gccaatttcg ctgggtcg cgtatgaaactc gggcagtcgt ttccaaactt tttggccgc
7801 cacgaccggc gcacccgggg ttggccggc gccgggttct tcagtcgcgc cggccaaacgg
7861 gggcgctgtc aggggcgtc cggcgacgac caacccgaag gacagcagcg cggagatcaa
30 7921 cggctcgccg cgccacatgg cggccatcgt ctcaccggcg aatacctgt acggcgca
7981 atgatcacac cttcggttct tcgccccgtc agcacttggc gccgctggc ggcgtgggtc
8041 cggcgattaa atacgcgtc acgtactcgta caatgcagct gtccgttgc aataccacc
8101 tgtgtgggt tccgtcgaaag gtcagcaacg aaccgcgaag ctgggtcgcc aggtcgaccc
8161 cggcgttgcgaa cggcgacggc gggcatggg tgggtggatc caccaccgtc ggcacttaggc
35 8221 cggcgccgaa gacggcatgg ggctgacttg tgggtggcac cggccagaac ggcgcagggtc
8281 ccagcggcgc atcaccgggt aacttccgt agctcatgaa cgggtgcgtc tcccgccgc

8341 ggcggcttc gtcatgacc ttgtcgat cggttaaccgg gggctgtatcg acgcaattga
8401 tcgccacccg cggtcaccgg gaatttgttgc agcggccgtg cgagtcccgaa cgcattgtaca
8461 tgctggccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agcccgctcg
8521 tcaagtgttg ccacagattc ggtagtaca ggcgcataat ggtgcccacg atggcgtcgc
8581 tataactcg cccgcgcgga tccttcgtgc ggcgcggccgt gtcgtatcgc gggtgtccg
8641 ggtcgaccaa cggatcgacc aggctgtggt agacccgtac ggctttggcc ggttcggcgc
8701 ccagcgggca gcccgcgttc ttggcgcagt cggcggcata gttgtgaac gcgtcctgga
8761 agcccttggc ctggcgcagc tccgcctcgat tggatcgatc attgggtcg acggcaccgt
8821 cgagaatcat tgccgcacc cgctgcggaa attcctcgatc atacgcggag cgcattccgg
8881 tgccgtacga gtagccccagg taggtcagct tgtagtgcgc caacgcgcgc cgaatggcat
8941 ccaggtcctt ggacgttgc accgtcccgat catggccag aaagtcttgc cccatcttgc
9001 ccacacagcg accgacgaat tgcttggatc cgttctcgat gtgcgcacaca cccctccggc
9061 ttagtcaac ctgcggctcg gcccgcagcc ggtagtgcgc ggcacatggag ttgcaccaga
9121 tcgcccggccg ggacgacgccc accccgcggg ggtcgaaccc aaccaggatcg aacccttgc
9181 gcaccgcgtt cggcaatgtc tggaaagacgc ccaaggccgc ctcgataaccg gattgcgg
9241 gtccaccggg atttatgacc agcgaaccga tcttgcgtcc cgtgcggga aagcgaatca
9301 ggcgcagcgc cggccacgtca ccatcggggc ggtcgatgc gaccggatc ggcgcgttgc
9361 cgcataacgc gccgcggggg atctttactt ggggttgc cgcacggcac ggtgtccact
9421 ccaccggctg gcccgcgttc ggctccgcata cgacgcgcg tcccccgcacc acgcggatgc
9481 agccacaag aaccaacgc acggcggcga ggcgcggccca gatcaacacgc atgcgcgcga
9541 tcttgcgcg gcgagacgc ctcatgcacc caatgtgcgc agacgcgacc cggatctcg
9601 gccagcggcc accgtcggcc gactaaccgg cgcgtcccg cagtcgtgc atgcgcgtatg
9661 gcgaaactcg cggccatccc ccatacgatcc ggtaaacagat cggggcaaga caccgacccg
9721 tcgaccggat cggcgcacggg cgcgtcgcc tcggcggatcg acaactgcga catcaggatcg
9781 ggcgtggcac cccgtccacg cggcatggt gcacccttggc catgcggccga gggcgatccc
9841 cgtatgcgcgc cacccttcg acgaacccat ctccacggc ggtcgccggc agcgcacgcga
9901 tggccgcac gatctccgag agtgcggcc gcccgcggg cgcacggcaac cccgtatgcgc
9961 gcaagtgcac atcgatgtga gttcaaggt tcagcgcact gctggcaagc ttttccgaa
10021 accgcggccct cgccttgcac tggagtcaga acgcgtcaccg cagccggtaa aaggcgtaac
10081 ccatgctcgat gcaaaatgc atggcgttag tggacgttgc cagacacacgc aactggcgtc
10141 caggccactg agccgcgtca tgcgcgtatgg tatgccgtatgg gggcccccgg ggcgcgtctga
10201 ggggaagaag tggcagactg tcagggtccg acgaacccgg ggaccctaacc gggccacgg
10261 gatgcaccccg accaccattt gggacagtgatgtca gagatctat gggccaaata
10321 ccccccggagg ctccggcccg cggaccaaga tccgcaccca ccacccatcg agatggaaagg
10381 cgcacggccca caagtggccatg atgcgtacccg cctacgacta ttgcacggcc cggatcttgc

10441 acgaggccgg catcccggtc ctgcgttcg gtgattcggc ggccaacgtc gtgtacggct
10501 acgacaccac cgtgccgatc tccatcgac agctgatccc gctggccgt ggcgtgggtc
10561 ggggtgcccc gcacgcactg gtcgtcgccg acctgcccgtt cggcagctac gaggcggggc
10621 ccaccgcccgc gttggccgccc gccacccggt tcctcaagga cggcggcgca catcggtca
10681 agctcgaggg cggtgagcgg gttggccgagc aaatcgccctg tctgaccgcg gcgggcattcc
10741 cggtgatggc acacatcgcc ttacccgc aaagcgtaa caccctggc ggcttccggg
10801 tgcagggccg cggcgacgccc gccgaacaaa ccatcgccga cgcgatcgcc gtgcggagaag
10861 ccggagcggtt tgccgtcgat atggagatgg tgcccgccga gttggccacc cagatcaccg
10921 gcaagcttac cattcccgacg gtcgggatcg gcgctggcc caactcgac ggccagggtt
10981 tggtagggca ggacatggcc gggtcageg ggcggcaagac cggccggctt gtcaaacgg
11041 atgcggatgt cggtggtgaa ctacggcgatc ctgcaatgca atacgccaag gaggtggccg
11101 gcggggattt ccccgctgac gaacacagtt tctgaccaag cggatcagc cggatcgccg
11161 ggcattgcgg tggcgccctg gatcgctcg acgcccggatt gccggccgg acgcgcac
11221 gggacccatc ggcgtcgctg tcgcccggat agcccggtt gagccagac attcgatgt
11281 cccaaacacca tccgcccacag cccaaatgtat gtcggactct atgcgtgcctt atccccgacc
11341 aaccaccacc cggcgacgc atcatgaccg gaggcgaaga tgccagttaga ggcggccaga
11401 ccagcgccgc atctggaggt cgagcgcaag ttgcacgtga tcgactcgac ggtgtcgccg
11461 tcgttcgagg gcatcgccgc ggtgggtcg gtcgacgactt cggcggccca gcagatcgac
11521 gcgggttact tcgacacacc gtcgcacgac ctggcgccga accagatcac ctgcggccgc
11581 cgcacccggcg ggcggcgcgc cggctggcat ctgaagctgc cggccggacc cggacaaggcg
11641 accgagatgc gaggcccgatc gtcggcatca ggcgacgctg tgccggccga gtttgtggat
11701 gtggtgctgg cgtatgtccg cggccaggcc gttcagccgg tgcggccggat cagactcac
11761 cgcggaaagcc agatctgtat cggcgccggg ggcgacgcgc tggcggaaatt ctgcacac
11821 gacgtcaccg catggcgcc cggggcattc cacggccgtg gtgcagccggaa caacggccct
11881 gccgaacagc agtggcgccga atgggaactg gaactggatca ccacggatgg gaccggccat
11941 accaagctac tggaccggct agccaaacggg ctgctcgatc cgggtggccgc acctggccgc
12001 cacggctcca aactggcgccg ggtgctcgatc ggcaccttc cgggtggatc gcccacccgc
12061 cccggccgc cggcgccatcc agtacaccgc ggcggccgg agcaagtcga cggatcgatc
12121 ctgtgggatc gggccgtcg ggcggacgccc tatgacccgg tgcaccatg gtcggatcgac
12181 acccgcaaga tccgcagctt gctgacggat tccaggat cgttgtggctt gaaggaaagt
12241 gcgtgggtca tcgatgaact gctgtggatc ggcgatgtcc tggcgttagc cggggacgc
12301 gaggtactcg gtgaccgcata ccagcgcgaa ctggacgcgc tggcgccgg gtcggatcg
12361 ggcgggggtgc gcgagccctt ggttagacggg ggcggccggc gataccagac cgggtcgcc
12421 cgatcactga tcgcattcg gtcgcagccg tacttccgtc tgctcgacgc tctagacgc
12481 ctgtgtccg aacgcggccca tgccacttctt ggggaggaat cggcaccggg aaccatcgat

12541 gccccctacc ggcgagtcgg caaagccgca aaagccgcaa agaccgccgg cgaccaggcg
12601 ggcgaccacc acccgacaga ggattgeac ctgatccgca agcgcgcgaa gogattacgc
12661 tacaccgcgg cggctactgg ggcggacaat gtgtcacaag aagccaagg catccagacg
5 12721 ttcttaggccc atcatcaaga cagcgtggtc agccggaaac atctgatcca gcaggccata
12781 gccgcgaaca ccgcggcga ggacaccctc acctacggtc tgctctacca acaggaagcc
12841 gacttggccg agcgctgccc ggagcagctt gaagccgcgc tgccaaact cgacaaggcg
12901 gtcgcggaaac cacgggattt agccgcggac gggccggacga tttggctgt aagccggatt
12961 ctgtccgcgc ccgcacacgc caagctaacg gcccacggc ggccaccatc catctggaca
10 13021 caccgttacc ggggcctcg agccgcctac ccgcaggctc gggcgagcaa ccctcaagcg
13081 cctgcgcggc cgcaacttcg gtgcggccctt ctggccctt ctgggggtgg ggtttgccta
13141 gccaccccgga tcacccggaa tgctggcgc ctcttaccgc accgttac ccttgccacc
13201 acgaggatgg cggtctgttt tctgtggcac ttcccgca gtcacctcg attggcgta
13261 gcaatcaccc tgctctgtga agtccggact ttctcgact cgacgtgaa ctcgtgaat
15 13321 ccacacaagg cctacgcgag ccgcggccgc ccagccaaact catccgcac gaccacgcta
13381 ccccgctggg cggtgtcgcc gccagttgtga ccgtggacg acacggctag tcggacagcc
13441 gatccggcgg gcaagtcccta tcgtggactg gtacacgggt gggacaaacg cgtcgactcc
13501 ggcgactggg acgccatcgc tgccgagggtc agcgagtcg gtggcgact gctacctcg
13561 ctgatcaccc cggcgaggc cggccggctg cgcaagctgt acgcccacga cggcctgtt
20 13621 cgctcgacgg tcgatatggc atccaagcgg tacggcgcc ggcagtatcg atattccat
13681 gccccctatc ccgagtgtatc gagcgtctca agcaggcgct gtatccaaa ctgctgcccga
13741 tagcgcgcaaa ctgggtggcc aaactgggccc gggagggcgc ctggccagac agccttgatg
13801 actgggtggc gagctgtcat gcccggccaa acaccgtatc cacagcgctg atgttgaagt
13861 acggcaccaa cgactggaaac gcccataacc aggatctcta cggcgagttt gtgttccgc
25 13921 tcgggttgt gatcaacctg agcgatccgg aaaccgacta caccggcgcc gagttcctgc
13981 ttgtcgaaca gcccgtcgc gccaatccc ggggtaccgc aatgcaactt ccgcaggac
14041 atggttatgt gtacacgacc cgtgtatggc cgggtggac tagccgtggc tggcggcat
14101 ctccagtgcgc ccatgggctt tcgactattc gttccggcga acgctatgcc atggggctga
14161 tcttcacga cgcagcctga ttgcacgcctaa tctatagata gcctgtctga ttccacaatc
30 14221 gcaccgacga tgccccatcg gcgtagaact cggcgatgtc cagcgatgcc agatcaagat
14281 gcaaccgata taggacgccc gaccggcat ccaacccgcag ccgcacaaac atttgtatcg
14341 cgtgacatg tgacaccacc agcaccgtcg cgcctcgta gccaacgtatc atccgatcac
14401 gtcggccgc aaccggccgc agcacgtcg cgaagetttc cccacccggg ggcgtgtac
14461 tgggtcctg cagccagcga cgggtcgact cgggatcgcc ttctggcc tccgcgaac
35 14521 tcagccccctc ccaggcgccg aagtccgtct cgcaccaggcgt gtcacgtacg accacgtcca
14581 gggccaggcgcgc tctggcgccg gtcacccgcgt tgcgtac cgcgtgtacg ggcgaggaga

14641 ccaccgcagc gatcccggcg cgccgcgcca gataccggc cgccgcacca acctggcgcc
 14701 acccccaccc tcgtcaaccccc gggttgcgcg gccccgaata gcggcggtgc tccgacagct
 14761 ccgtctgccc gtggcgcaac aaaagtatgc gggtgggtgt accgcgggccc cccgtccagc
 5 14821 cgggagatgt cggtgactcg gtcgcaacga tttggcagg atccgcattcc gccgcagccg
 14881 attgcgcggc ggcgtccatc gcgtcatgg ccaaccggc tgcatacgtg ttccggcac
 14941 gccaaccca ctctgtatgc atcctgcgaa actgggacgc caacgcctga gcctggacat
 15001 agagcttcag cagatccggg tgcttgacct tccaccggcc ggacatctgc tccaccacca
 15061 gcttggatgc catcagcacc gcggcctcg tggcacccat tttcacggcg tcgtccaaac
 10 15121 cggctatcag gccgcggtat tcggcgacgt tggcgatgc cccggccatc gcctgcgttg
 15181 actcggccag cacgggtggag tggatcggcg tccacaccac cgcggcgat cccggccggc
 15241 cgggattgcc cccgcgtcccg cccgtccgtt cgtacacaac tttactctt caaatccttc
 15301 gagccgcaac aagatcgctc cgcattccgg gcagcgcacc acttcattcc cggcggccgc
 15361 cggatctgg gcaagctcg cccggccat ctcgatccgg caggcaccac atcgatgacc
 15 15421 ttgcaaccgc cccggccctg gcccgcctcc gggccgtgt cttagctaga gccccgcaag
 15481 ctccggatca agtgtcgccg tcagcatgtc gcgttgcgtt gaatgttggt gccgggttg
 15541 gtcgatttcg gcaagtgcct cgtccaaaggc ctgcggcg gccggccaggc cggccggcaa
 15601 cgttggagc gcccggact cggcggtctg ttgagctgc agtcctcg cggcggtccag
 15661 caccccgccggc agggcatctt cccaaactggc ttgacggcg tgcacgtgt cggatctgt
 20 15721 ctgcagatca gccaattgtt tggcgatccgt tgacccggaa gtgagcaacg accgggtcccg
 15781 gtcggccacgc ttacgcaccg catcgatctc cggatctt cggacacaccc ggcgtccaa
 15841 gtcctccggc gcgatcgca gggccggccat cctgtcggtt gccggcgat gctccggctg
 15901 cacctgctgg taagccggccc gctgcggcag atggtagcc cgtacgcgca tccgggtcag
 15961 ctcagcatcc agttcgcca attccagtag cggccgttgc tgcacactc cgggttcat
 16021 gcccgtatcc tcggatccgtt gtcgtcgagg ttccacgggtt cggatcgat ggtgcacaca
 25 16081 cgcacccggca gcgacgcggc gaaatgagac cggacacactt cggccggctg gcccgcaccac
 16141 gggaaattcgc ttggccaaatgc cggacgtcg atcaggggccat cttgcacgc tcggcaatgc
 16201 tcgtcggtcg gatgtatcg cagatcgccg gtaacgtacg cttgcacgtc cccggccggcc
 16261 acggatggca gcaacggatc cccggccggc cccggccggc cggacccgcga caccagcagg
 16321 tcggatccc cggccggcg caccggcgtc cggatcgatc gcaacgcggc ctccagacgg
 30 16381 gcaacaaagg tggcgatccgtt tggcgatccgtt ggcgtatgc caatccggcc taaccggctg
 16441 cccggccggc gttggatccgt cggatcgatc tggcgatccgtt ggcgtatgc caatccggcc
 16501 gcgacgtcg cccggccggc cccggccggc gttggatccgtt ggcgtatgc caatccggcc
 16561 tcctcgatcc cccggccggc ggtaccgcgt cccggccggc gttggatccgtt ggcgtatgc
 35 16621 gccaggaact gcccggatcc cggacactc cggatcgatc tggcgatccgtt ggcgtatgc
 16681 cccggccggc cccggccggc ggtaccgcgt cccggccggc gttggatccgtt ggcgtatgc

16741 atgacccact tgtcgagatc ggccgctccg ggcacccgggt cgagaacggc gtcgacggtc
16801 agaccaacag cgtgtgccag cgcgctggac acaccggcg acgcccagtc ggcgttggtg
16861 tgcgcggtaa acaacgagcg accggctccgg atcaggcggt gcaccagcac accctttggc
5 16921 gtgtggccg cgaccgtatc gaccccacgc agtaacaacg ggtggtcac caatagcagt
16981 ccggcctggg gaacctggtc caccaccgccc ggcgtcgctt ccaccgcaac ggtcaccgaa
17041 tccaccacgt cgtcggggtc gccgcacacc agacccacccg aatcccacga ctgggcaagc
17101 cgcggcgggt aggccctggc cagcacgtcg atgacatcg ccagccgcac actcatcgcc
17161 gtcctccacg ctttgcac tcggcgatcg ccgcccaccag cacgggcccc tccgggccc
10 17221 ccggccggcc caggtaccgc ggcgtccaggc cgaccaagggt gtacccggcgg cgcaccggcaa
17281 ttccattgtct ctgcaaatacg ttccgtaaatc cgtcagcatc ggcgtatgtt aacagtacga
17341 aaggggccgc accatcgacc acctcgccac ccaccgatct cagtcggcc accatctccg
17401 cgcgcagcgc cgtcaaccgc accgcattcg ctgcggcagc ggcgaccgccc cggggggcgc
17461 agcaaggcgc gatggccgtc agttgcaatg ttcccaacgg ccagtgcgtc cgtgcacgg
15 17521 tcaaccgagc cagcatgtt ggcgagccga ggcgttagcc cacccgcaat cggccagcg
17581 accacgtttt cgtcaagcta cggagcacca gcacatcggtt cagcgactca tcggccaacg
17641 attgcggctc gccgggaacc caatcagcga acgcctcgatc gaccaccagg atgcgtcccg
17701 gccggcgtaa ctgcgagcagc tgctcgccga ggtgcagcac cgagggtgggg ttggcgat
17761 taccacgac gacaaggctcg cgcgtcgatc gcacgtgcgc ggtgtccagc acgaacggcg
20 17821 gctttaggac aacatggtgc gccgtgattc cggcagcgctt caaggctatg gccggctcg
17881 tgaacgcggg cacgacgattt gctgcccgc cccggacttag gttgtgcagc aatgcgaatc
17941 cctccggccgc cccgacgagc gggagcactt cgtcacgggt tctggcatga cgttcagcga
18001 ccgcgtcttg cgcccggtgc acatcgatc tgctcgatc gccggccagc tccggcagca
18061 ggcggcggag ctgcccggacc aaccattccg gggccgggtc atggcggacg ttgacggcga
18121 agtccagcac gcccggcgcg acatcgatc caccgtgtt ggcgcggcgc gcaaggcggc
25 18181 tagtgtctat actcgccaca ggcgtaaaca gtatggcc ggtgtgcggg ccaagaatcc
18241 agagcaccgc cgacgcgttg tctacgcggc gacaaccgcg acatcacagg cagctaacag
18301 ggcgtcgccgc gtgtatcgatc tcaggccaaag cagctgtgc tggcgatga gcacacggc
18361 gaatggatgt cgatggatgt cggaaagctc tgccgtgcgc agtgtgtgcg tggtaactg
18421 acagcggcga cgtgcgcag cggcgcattt gatgggcac gtaagaagcc gatggctcg
30 18481 gccggcggag ctggccgagg cggtagtta tcgcgtatctt ccaggactg gccggccgaca
18541 agagaatgtt gtgcggacg tcctgaacaa tcggccgtt ttcgttgcgc gcatccgc
18601 ccaaacgtgg gtgtcgatgc ggtatcgatc caccgggtt gacgttcgcg cacgtcgatc
18661 gacaacggag cgtccaaatc gtccggcgcg cggatcacgc catggtcaat gcctaaccgc
35 18721 cgagtctat gaggatgcag cggcacaaggc tttgttccgc gtcgcgcgc gccggcaatc
18781 tcaacctctg cccggcgtag acgagccgca gcagctcgga caggcgtgc ttcgcctcg

18841 gaacgccgac ccgcgtcgca ggcccaga cttcgctc gaccacctgc tcaccaaact
18901 tcgcgatcat cgccgtatac cacagcgcca acggtagcg gtttgccta ccgcgtcgca
18961 aacgacaatg ggatcggtac cgacacgacc gcgagcggga ccaattgccc gcctctcca
5 19021 cgcgcgcgcg cacggcgccgc atcgctcgccg ggtgaatcgcc cgacgttgtt gatcttcgat
19081 ctggacggca cgctgaccga ctggcgccgc ggaatcgat ccagcttccg acacgcgc
19141 aaccacatcg gtgcggcagt acccgaaggc gacctggcca ctcacatcg tggcccgccc
19201 atgcgtgaga cgctgcgcgc catgggctc ggcgaatcgcc cgaggaggc gatcgtagcc
19261 taccggcccg actacagcgcc cgcgggttgg gcgatgaaca gcttgcga cggatcg
10 19321 ccgcgtgtgg ccgcgtcgcc accggccgtt gtccggctgg cgctgcac ccacaggca
19381 gagccgaccg cacggcgaat cctgcgcac ttggaaattt agcagcactt cgaggcatc
19441 gccccggcga gcaccgttgg ctgcgaggc agcaaggc acgtgtggc ccacgcgc
19501 ggcgcagctgc ggccgcgtacc cgagcggttgg gtgtggctcg ggcgcacgc ccacgcgc
19561 gacggggccgg cgcgcacgg catgcacacg gtgggttgg gctggggcta cggccgc
15 19621 gactttatcg acaagaccc caccaccgtc gtgcgtcgat ccgcacat tgacgagct
19681 agggaggcgcc taggtgtctg atccgcgtca cgtcacattt gttgtacgg gcaacatctg
19741 ccgggtcgcca atggcgaga agatgttgc ccaacagtt cgccaccgtg gcctgggtga
19801 cgcgggtcgca gtgaccagtg cgggcacgg gaaactggcat gtggcagtt ggcgcacga
19861 gccccggcggcc ggggtgttgc gagccacgg ctaccctacc gaccaccggg cgcacaagt
19921 cggcaccgaa cacctggccgg cagaccgtt ggtggcttgc gaccgcacc acgctcg
20 19981 gttgcggcag ctggcggtcg aagccgcggc ggtacggat ctgcgtcat tcgacccac
20041 ctggaaacc catgcgtcg atgtcgagga tccctactat ggcgtactt ccgacttgc
20101 ggagggttcc ggcgtcatcg aatccgcctt gcccggcttgc caccgttgg tcgacgaac
20161 tctcgccggg aacggaccga gtgtatggcc cgcctagctt tcctgtcg gcccggctt
20221 ctggcggttgg ccctgggtcg ggtgcgttc acctaccgtt gctcgccgg
25 20281 tggcagctgg gcaagaatgc caaaacgtca cgagagaacc agcagatcg gtattccctc
20341 gacaccccgcc cggttccgtt gaaaaccctt ctaccacaggc aggatcgcc ggcgcggac
20401 ggcgcgttggc gcccgggtgac ggcaaccggc cagttccgc cggacgttgc ggtgtggcc
20461 cgcgttggc tgggtggaggc ggaccaggcg ttggagggttgg tggccctt cgtggcgac
30 20521 ggcggaccaa ccgtcttggt cgaccgttgg tacgtcgccccc ggggttggg ctcgcacgt
20581 ccaccgttcc cccgccttgc ggtgcagacg gtgaccatca ccgcgcggct gctgtactt
20641 gaaccggcgcg tggccggccaa agaccattt gtcagagacg gttccagca ggtgtattcg
20701 atcaataccg gacaggttgc cgcgttgcacc ggagttccagc tggctgggttgc ctatctgc
20761 ttgtatcgaa accaaccggc cgggttgcgc gttgttgc gttccatctt agatcccg
35 20821 ccgttcttgc cctatggcat ccaatggatc tcgttgcgc ttctggcacc gatcggttgc
20881 ggctatttgc cctacggccga gatccggccgg cggccggccggg aaaaagcgccc gtcggccacca

20941 ccggacaaggc caatgacggt cgagcagaaa ctgcgtgacc gctacggccg cggcggtaa
21001 accaacatca cggccaatac cgcagcccc gcctggacca cccgcgacag caccacggcg
21061 cggcgacat cggccaccc gggcgaccgg cgctgcccc aggtggccg gatctgcaac
5 21121 tcatggtgtt accgggtggg cccaccac cgcacgtaa gcccggcagc aaacggcc
21181 tcgacgacac cggcggtggg gctggatgg cggcgccgt cgccgcgcca ggccgtacc
21241 gcaccgggg gcgacccacc gaccacggc ggcagatca ccaccagcac cgccgtcgcc
21301 cgtgcgccaa catagttgc ccagtcatcc aatctgtct cagcccaacc gaatcgaga
21361 taacgcggcg agcggtagcc gatcatcgag tccagggtgt tgatggcagc atatcccagc
10 21421 accgcaggca cggcgtcga agccgcccac agcagcggca ccacctggc gtcggggtg
21481 ttgcggcca cgcactccag cgcggcacgc gtcaggcccg ggccgcccag ctggccggg
21541 tcacgcccgc acagcgacgg cagcaggcgt cgccggcct cgacatcgac ggcgtccaaac
21601 aggtccgata tctggcggcc ggtgcgcgcc agcgaagtgc cgccagcgc tgccaggtg
21661 gccgtcgccgg tggccgcccac gggccaggac ctggccggta gccgctgcag tgccgcgc
15 21721 agcaagccca cgcgcggcgcac cagcaggccg acgtgtaccg caccggcgcac cggccgtca
21781 cggtaggtga tctgctccag ctggccggcc gcccgaccga acagggccac cggatgaccc
21841 cgttgggtt cgccgaacac gacgtcgagc aggccggcga tcagcagcgc gacggccctg
21901 gtctgccagg tcgatgcaaa cactccggca gcgtcgacaca cgtggtctac gctcagctat
21961 ttatgaccc atacggcagc tatccacgt gaagcggcca gctacccggg ttgccgaccc
20 22021 gttgaacccg gcggaatgt tggccggc agcgaatgtc atcatgcgc tgccagtgc
22081 ggggtcgccg tggaaagccc ggtggacagc ggcaacgtct acaagcatcc
22141 gtcaagcgg gcccggacca cggcaccta ctggccggtg ggcaccatcg ggacggaaatc
22201 cgaccgagcg ctgatccggg gtggccgtga cgtcgccac cggcagggttc ggtcgacggc
22261 ctcgagccca gtgtcctata acgcctcga cccgaagtgc cagctgtggg tggccggcgt
25 22321 tctgtaccgc tactcggtt accagcacga gtttctgtac gcccactcg aagatgccac
22381 cggcgcgcgc gtctaccaag acgcacaaacg gtttagggacc acgctgcagg tgccggaggg
22441 gatgtggccg cggaccggg tcgcgtcga cgactactgg aagcgctcgc ttgtatggct
22501 gcagatcgac ggcgggtgc gcgacatct tcgcgggggtg gcctcggtag cgttctccc
22561 gtggccgtt cgcgggtgg cggggccgtt caacctgtt ggcacgacgg gattctggc
30 22621 accggagttc cgcgcgtatga tgcagctgga gtggtcacag gcccaggcgc gtcgttcga
22681 gtggttactt tccgtgtac gtttagccga cggcgtgatt ccgcacatggg cctggatctt
22741 cgttaccag ctttacttgt gggacatcg gtttcgcgcc cgacacggcc gccgaatcg
22801 ctgatagagc cggccgagt gtgagccgtga cagcccgaca cggcggcgt gtgtcgcc
22861 gccagggtca cgcgtggcga tctagagccg cggaaaacct acttctgggt tgccctccga
35 22921 atcaacgtgc tgatctgctc gagcagctca cgcacatcg ggcacatcgcc atccaccgc
22981 gcatacaggt cggccttggt cggccggcgc tggccgcacg tcattggccg caccggcggt

23041 gctgtctgtc ggcggcgct gtcgcttga aaccaggc gtcacccac gaccacgaca
23101 ctgccatatc cggcgccccg ccgacaacga agcacagta gccggggc gggacggga
23161 tcgaaccgcc gaccgcttgt gtgtaaaacc agagctctac cgctgagcta cgcgccatg
5 23221 accgcccag gctacacgccc ttggccaa gcacccaaaa ccttaggccc taagcgccgc
23281 cagagcgtcg gtccacagcc gctgatcgca aacttcaccc ggctgcttca tctcgccgaa
23341 ccgaatgatc cctgaccgat cgaccacaaa ggtgccccgg ttagcgatgc cggctgctc
23401 gttaagacg ccgttaggcct gactgaccgc gccgtgtggc cagaagtccg acaacagcgg
23461 aaacgtaat ccgctctgca tcgcccagat ctgtgagtg ggtggggc ccaccgaaat
10 23521 cgtagcgcg ggcgtgtcg cgttctaaa ctggcagg tgatcagca actggccag
23581 ctgcacctgg cagatgcccc tgaacgccaa cggaaagaac accaacagca cgttcttgc
23641 accccggtag ccgcccagg tgacaagctg ctgattctgg tcgccaacg tgaagtcagg
23701 ggcggggct ccgacgttca gcatcagcgc ttgcagccc gcatcgcc ctgtaccaat
23761 ctgtggcgc tccagttgcc cagattgacc gacgaggctg gcatcagccc agctgtggc
15 23821 gccgcctcgg caatctggc gggcaataca tggccggct ggccgttgc gggcgtacc
23881 acccaaataca caccgtccctc ggcgagcggg ccgatcgat ccatcagggt gtccaccaaa
23941 tcgcgtcgc catcacgcca ccacaacagg acgacatcga tgacctcgatc ggtgtttca
24001 tcgagcaact ctcccccga cgttctcg atggccgcgc ggtatcgatc gtcgggtct
24061 tcgtcccagc cccattctg gataagttgg tctcggttga tgcccaattt gccccgttag
20 24121 ttcgaggcgt gatccgcgc gaccacgtg gaaccttctt cagtccgc gggcatgtg
24181 cacaccgtcg cgtatggcat tatcgatcgca cagccagaac cggccaccc gcccgcctca
24241 gaaggcggcc acgcacattt tcaatcgatc tgcgttgc gatcaacccg
24301 ccgggttaat tccgctgtcg acgcgtgcgc accgatggca ttgcacccg cggggccgc
24361 gtcgacatat gcggtgagcg catcccccag ttgcggac agcgcggcgc tcagactgcc
24421 tgagaccgtc gaggcactgt tggatcgatc gtcgtggcc ggacccctgg tcggccgggt
25 24481 ttgcggcc tggatcgatc cggccacgtt ggcgttccacc ttgtcgatgg cgtcccttgc
24541 gttggccgc acgcgtcgc acgagggtcg aatcgatcgatc gtcgtggcc attttggcg
24601 ctgcgactcc cggatcgatc acgtcgccgc cgaaggccac accgacgcgg acaccgacga
24661 gccgttaggccc ggtgcgacgt tggatcgatc catggccgtt gtcgtggcc cagtggata
24721 tccgacgatc cccatcgatc gcaacgtcgatc gtcgtggcc ggacccctgg tcggccgggt
30 24781 gagctcccc cccatcgatc gaggcactgt ggcgttccacc atgaggacgg catgtgaggt
24841 tacctggatcgatc cggccgcacc ggcgtggcc tggatcgatc gtcgtggcc gatccgacgc
24901 gagatcgatc atccggccgc gacgcgggtt ggcgttccacc gggggacgc acaccgacga
24961 gcaacgtcgatc ggaaggccgc cacatcgatc agtgcgttgc tgaccaccga ttgcggccgc
35 25021 cacatcgatc cccatcgatc aaacagcgatc agcgaacccg accgagttcg ggtgcgttgc
25081 gagggtgtgg cgtcgatcc gcccacattt gatccgagg agaccccgatc gtcgtggag

25141 tccttgaca cgctgctgca acgctgcggc ccgtcgccgg cccgctacct gatgtgcgg
25201 ctgctagagc gggccggcga gcagcggtg gccatccgg cattgacgac taccgactat
25261 gtcaaaccca tcccgaccga gctggagccg tgggtccccg ggcacgaaga cgtcgaacgt
5 25321 cgttatcgag cgtggatcag atggaatgcg gccatcatgg tgcaccgtgc gcaacgaccg
25381 ggtgtggcgc tgggtggcca tatctcgacc tacgcgtcg cgcggcgct ctatgaggtc
25441 ggittcaacc acttctccg cggcaagtgc cacccggcgc gggcgatca ggtgttcac
25501 cagggccacg ctccccggg aatctacgcg cgcgcctcc tcgaaggcgc gttgaccgccc
25561 gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcgggtt gccgtcstat
10 25621 cgcacccgc ggctcatgcc cgacttctgg gaattccca cgtgtcgat gggttgggc
25681 ccgctcaacg ccatctacca ggcacggtc aaccactatc tgcatgaccg cggatcaaa
25741 gacacccctcg atcaacacgt gtggtgttt ttgggcgacg gcgagatgga cgaacccgag
25801 agccgtggc tggcccacgt cggcgctgc gaaggcttgg acaacttgac ctctgtatc
25861 aactgcaatc tgcagcgact cgacggcccg gtgcgggca acggcaagat catccaggag
15 25921 ctggagtcgt tctccgggg tgccggctgg aacgtcatca aggtggtgtg gggccgcgaa
25981 tgggtatcccc tgctgcacgc cgaccgcac ggtgcgtgg tgaatttaat gaatacaaca
26041 cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtgcg tgaccactc
26101 ttccggccgcg acccacgcac caaggcgctg gtggagaaca tgagcgacca ggatatctgg
26161 aacctcaaac ggggcggcca cgattaccgc aaggttacg ccgcctaccg cggccgcgtc
20 26221 gaccacaagg gacagccgac ggtgatctg gcaagacca tcaaaggcta cgcgcgtggc
26281 aagcatttcg aaggacgcaa tgccacccac cagataaaaa aactgaccct ggaagacctt
26341 aaggagtttc gtgacacgcg cggattccg gtcagcgacg cccagcttga agagaatccg
26401 tacctgccgc cctactacca cccggcctc aacgccccgg agattgtta catgtcgac
26461 cggcgccggg ccctcggggg ctgttccc gagcgcagga ccaagtccaa agcgctgacc
26521 ctggccgggtc ggcacatcta cgcgcgcgt aaaaagggt ctggcacca ggaggtggcc
25 26581 accaccatgg cgacgggtcg cacgtcaaa gaagtgtgc ggcacaagca gatcggcccg
26641 cggatagtcc cgatcattcc cgacgaggcc cgcacccctcg ggtggactc ctgggtcccg
26701 tcgctaaaga tctataaccg caatggccag ctgtataccg cgggtgacgc cgaccctgatg
26761 ctggcctaca aggagagcga agtcggcag atcctgcacg agggcatcaa cgaagccggg
26821 tcgggtggct cgttcatgc ggccggcacc tcgtatgcgac cgcacaacga accgatgatc
30 26881 cccatttaca tcttctactc gatgtcgcc ttccagcgca cggcgatag ctctggcc
26941 gccggccgacc agatggctcg agggttcgtg ctggggcca cccggccggcgc caccaccctg
27001 accgggtgagg gcctgcaaca cggcgacggt cactcggtc tgctggccgc caccacccg
27061 gccgtgggtt cctacgaccc ggcctcgcc tacgaaatcg cctacatcgat gaaagcgga
35 27121 ctggccagga tgtgcgggga gaacccggag aacatcttct tctacatcac cgtctacaac
27181 gagccgtacg tgccggcc ggagccggag aacttcgatc ccgaggcgt gtcgtgggtt

27241 atctaccgct atcacgcggc caccgagcaa cgcaccaaca aggcgcatat cctggcctcc
27301 gggtagcga tgcccgccggc gctgcgggca gcacagatgc tggccgccga gtggatgtc
27361 gccgcccacg tgtggtcggt gaccagtgg ggcgagctaa accgcgacgg ggtggccatc
5 27421 gagaccgaga agctccgcca ccccgatcgg ccggcgggca tgccctacgt gacgagagcg
27481 ctggagaatg ctcggggccc ggtgatcgcg gtgtcgact ggatgcgcgc ggtccccgag
27541 cagatccgac cgtgggtgcc gggcacatac ctcacgttg gcaccgacgg gttcggttt
27601 tcggacactc ggcccgcgc tcggcgtac ttcaacaccg acgccaatc ccagggtggc
27661 gcggtttgg aggcgttggc gggcgacggc gagatcgacc catcggtgcc ggtcgccggc
10 27721 gcccgcactg accggatcga cgacgtggc gtcgcggcc agcagaccac ggatccgggt
27781 cccggggcct aacgcccggc agccgaccgc ctggccga atctccaga aatctggcgt
27841 agcttttagg agtgaacgac aatcagttgg ctccagttgc ccgcccggagg tcggcgtcgc
27901 aactgctgga cactgtgcc gattcgctc tgcggcggtt gaagcagtac tcggccggc
27961 tggccaccga ggcagttcg gccatgcaag aacgggtgcc gtttcgccc gacctagaag
15 28021 cgtcccacgc cgccagegtg gegetggtgg tgcagacggc cgtggtaac ttgtogaat
28081 ggatgcacga cccgcacagt gacgtcgct ataccgcga ggcattcgag ctggcccc
28141 agatctgac ggcacggatc ggcgtgcgc agaccgtgga catggtgcgg gtcacccatgg
28201 agtcttcga agaagtcgtg cccctgtcgc cccgttccga agagcagtgg accgcctca
28261 cggggccat ttgaaatac agccgcgacc tggcattcgc ccggccacg gcctacggc
20 28321 atcggccga ggcacgaggc acctggaca gcccggatgga gcccagcgtg gtggacgg
28381 tggtaacgcg cgcacccggt cccgagctgc tgtccgggc ggcgcgcgtg aattgggaca
28441 ccaccgcgc ggcgaccgta ctggggaa ctccggcgc cggccaaat ggctccaaca
28501 ggcacggcga cagcgagcgg gccagccagg atgtccgcga caccgcggct cggccacgg
28561 ggcgtgcgt gaccgacgtg cacggcacct ggctggtggc gatcgctcc ggcagctgt
28621 cggcaaccga gaagttcctc aaagacctgc tggcagcatt cggcgcgc cccgtggta
25 28681 tggccccac ggcgcctatc ctgaccgcgg cgcacccgc cgttagcgc ggcgttcgc
28741 ggatgaacgc cgtcgccggc tggcgccggag cgcgcggcc cgtgtggct agggacttt
28801 tggccgaacgc cggccgtatc ggcgcacgc cggcgatcgt ggcctgcatt accgacgtga
28861 tggcccccct agccgatgcc ggaccgacgc tcatcgagac gctagacgca tatctggatt
28921 tggcgccgc gattgaagct tggccagaa agttgtcg tcatccaaac acagtgcgg
30 28981 accggctcaa gcggtaccc gacttcaccc ggcgcgtcc caccgcgc cgcgtgcct
29041 atgccttcg ggtggccgc accgtggc aactcaacta tccgacgcgc cactgaagca
29101 tcgacagcaa tggcggtca tagattccct cggccgtcag agggggtcca gcagggcc
29161 cggaaagata ccaggggcgc cgtcgacgg aaagtgtacc agacaacagg tcgcccgg
35 29221 atctaaaaaa catagcttac agggccgtt tgggttat atacaaaaac ctaagacgag
29281 gttcataatc tggtaacccg cgcggccatc tttcacagt gttctcttag acacgtgatt

29341 gcgttgctcg cacccggaca gggttcgcaa accgaggaaatgttgcgcgtggc
 29401 ctgcccggcg cagcggacca gatcgcggcg tggtcgaaag ccgctgatct agatcttgc
 29461 cggctggca ccaccgcctc gaccgaggag atcaccgaca ccgcggcgcgc ccagccattg
 5 29521 atcgtcgccg cgactctgct ggcccaccag gaactggcgc gccgatgcgt gctcgccggc
 29581 aaggacgtca tcgtggccgg ccactccgtc ggcgaaatcg cggcctacgc aatcgccgg
 29641 gtgatagccg ccgacgacgc cgtcgcgtc gccgccaccc gccgcgcga gatggccaag
 29701 gcctgcgcca cccggccac cggcaigtct gccgtgcgtc gccgcgcga gaccgaggtg
 29761 ctgagtcgccc tcgagcagct cgacttggc ccggcaaacc gcaacgcgc cggccagatc
 10 29821 gtcgctggcg gccggctgac cgcgttggag aagctgcgc aagaccgcgc gccaaggcg
 29881 cgggtgcgtg cactgggtgt cgccggagcg ttccacaccg agttcatggc gcccgcactt
 29941 gacggcttgc cggccgcgc gccaacatc gcaaccgcgc accccaccgc cacgctgcgt
 30001 tccaaccgcgc acgggaagcc ggtgacatcc gccgcgcgg cgatggacac cctggctcc
 30061 cagctcaccc aaccggtgcg atgggacctg tgacccgcga cgctgcgcga acacacagtc
 15 30121 acggcgatcg tggagttccc ccccgccggc acgcttagcg gtatgccaa acgcgaactt
 30181 cgggggggttc cggcacgcgc cgtaagtca cccgcagacc tggacgagct gccaacaccta
 30241 taaccgcgga ctggccaga acaaccacat acccgtcagt tcgatttgcata cacaacatcat
 30301 tacgaaggga agcatgcgt gcctgtcact caggaagaaa tcattgcgg tatcgccgag
 30361 atcatcgaag aggtAACCGG ttcgagccg tccgagatca ccccgagaa gtcgtcg
 20 30421 gacgacctgg acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag
 30481 tacggcgta agatccccga cgaggaccc tcgggtctgc gtaccgtcgg tgacgtgtc
 30541 gcctacatcc agaagctcgaa ggaagaaaac ccggaggcgg ctcaggcggt ggcgcgcgaag
 30601 attgagtcgg agaaccgcga tgccgtgcc aacgttcagg cgaggctga gcccggatcc
 30661 aagttagtca gccttcacc gctaattggcg gttcccccag cgttgtggt accgcgcgtca
 25 30721 cagcgacgac gtcgatctcg ccggacatcg agagcacgtg gaagggtctg ttggccggcg
 30781 agagcggcat ccacgcactc gaagacgagt tcgtcaccaa gtggatcta gcggtcaaga
 30841 tcggcggtca cctcaaggat ccggcgcaca gccacatggg ccgactcgac atgcgcacgca
 30901 tgtcgtagt ccagcggatg ggcaagtgc tggcggaca gctatgggag tccgcggca
 30961 gcccggaggt cgatccagac ccgtcgccg ttgttgcgg caccggctca ggtggagccg
 30 31021 agaggattgt cgagagctac gacctgtatcg atgcggcgg ccccccggaaatgttccgc
 31081 tgccgttca gatgatcatcg cccaaacgggt ccgcggcggt gatcggtctg cagcttgggg
 31141 cccgcgcgg ggtgtatgacc ccgggtcggt cctgttcgtc gggctcgaa ggcgcgc
 31201 acgcgtggcg tcagatcgatcg atggcgcacg ccgacgtcgc cgtctgcggc ggtgtcgaaag
 31261 gaccatcgaa ggcgcgtccc atcgccgggt tctccatgt gcccggcatcg accgcgc
 35 31321 acgacgagcc tgagcggccccc tcccgccgt tcgacaagga ccgcgcacggc ttgtgttcg
 31381 gcgaggccgg tgcgtatcgatcg ctcatcgaga cggaggagca gccaacagcc cgtggccca

31441 agccgttggc ccgattgtcg ggtgccggta tcacctcgga cgccttcat atggtggcgc
31501 ccgcggccga tgggttcgt gccggtaggg cgatgactcg ctcgtggag ctggccgggt
31561 tgtcgccggc ggacatcgac cacgtcaacg cgcacggcac ggcgacgcct atcggcgacg
5 31621 ccgcggaggc caacgccatc cgcgtcgccg gtgtgatca ggccgcgtg tacgcgcccga
31681 agtctgcgt gggccactcg atcggcgccg tcgggtcgct cgagtcggtg ctcacgggtgc
31741 tgacgctgcg cgacggcgtc atccgcgcga ccctgaacta cgagacaccc gatcccgaga
31801 tcgaccttga cgtcgccgc ggcgaaccgc gctatggcga ttaccgctac gcagtcaaca
31861 actcggttcgg gttcggccggc cacaatgtgg cgcttgcctt cggcggtac tgaagcacga
10 31921 catcgccgggt cgcgaggccc gaggtggggg tcccccgct tgccccggc agtcggaccg
31981 atatgaaagg aacgttcgca agaccaatga cggagctggt taccggaaa gcctttccct
32041 acgttagtcgt caccggcatac gccatgacga cgcgcgtcgc gaccgacgcg gagactacgt
32101 ggaagttgtt gctggaccgc caaagcggga tccgtacgct cgatgaccca ttctgcgagg
32161 agttcgacct gccagttcgc atcggcggac atctgttga ggaattcgac caccagctga
15 32221 cgcggatcga actgcgcgg atgggatacc tgcagcggat gtccaccgtg ctgagccggc
32281 gcctgtggga aaatgcggc tcacccgagg tggacaccaa tctgttgcgt gtgtccatcg
32341 gcaccggcct gggttcggcc gaggaactgg tcttcagttt cgcgcgtcgc
32401 gaatgaaggc ggtctcgccg ctgaccgtgc agaagtacat gcccacggg gcccggcgg
32461 cggtcgggtt ggaacggcac gccaaggccc ggggtatgac gccggatcg cggtgcgcac
20 32521 cccgcggcga ggccatcgcc cgtcggtggc agcagattgt gctgggagag gccgatggc
32581 ccatctgcgg cggcggtggag accaggatcg aagcggtgcc catcgccggg ttgcgtcaga
32641 tgcgcacatcgatgtccacc aacaacgcacg accccgcgg tgcacgcgc ccaattcgaca
32701 gggaccgcga cggctttgtt ttcggcgagg gcccgcctt tctgttgcgt gagaccgagg
32761 agcacgccaa ggcacgtggc gccaacatcc tggccggat catggcgcc agcatcacct
25 32821 ccgatggctt ccacatggtg gccccggacc ccaacggggaa acgcgcggg catgcgatta
32881 cgcggcgat tcagctggcg ggcctcgccc cggcgacat cgaccacgtc aatgcgcacg
32941 ccacccggcac ccaggtcgcc gacctggccg aaggcagggc catcaacaac gcctggcg
33001 gcaacccgacc ggcgggttac gcccccaagt ctgcctcgcc ccactcggtt ggcgggttgc
33061 ggcgggtcgatcgatctt acgggtcgatcg cttgcgcga tcagggtatcg ccggcgacac
33121 tgaatctgtt aaacctcgat cccgagatcg attggacgt ggtgggggtt gaaccgcac
30 33181 cgggcaatta ccggatcgatcg atcaataact cttgcgttgcgtt cggcgccac aacgtggcaa
33241 tcgccttcgg acggactaa accccagcgat tacgcgcacag gagacactcg atgacaatca
33301 tggcccccga ggcgggttgcg gagtcgtcg accccgcga tccgtgttgcgtt cggctgagca
33361 actcttcgatcg acggcgac gttggatcg tgcacgcac gtcggatcg ccggcgacac
33421 ccgcggcgccg caccgtcaac ggtgtgcgcac ccatcggtt ctgcaccgc ggcaccgtga
35 33481 tggccggcgccg catggcgatcg gaggggttgcg cgcacatcgatcg caacgcctac gacactgcga

33541 tcgaagacca gagtcggatct gtgggcacgt ggcattcggg tgggtcccggt ctggctgaag
33601 gtgtcgcccc gctgcacgcg gtaggccagg tggtcaagc catgatccgc ggttccggct
33661 acatcccgca gatctcggtg gtcgtcggtt tcgcccggg cggccggcc tacggaccgg
5 33721 cgttgaccga cgtcgctgc atggcgccgg aaagccgggt gttcgtaacc gggcccgacg
33781 tgggtcgccag cgtaaccggc gaggacgtcg acatggcctc gtcgggtggg ccggagaccc
33841 accacaagaa gtccgggggt tgccacatcg tcgcccacga cgaactcgat gcctacgacc
33901 gtgggcggccg gtgggtcgga ttgttctgccc agcaggggca ttgcgtcgcc agcaaggccg
33961 aggccggta caccgacatac cacgcgtgc tgccgaaatc ctcgcgacgt gcctacgacg
10 34021 tgcgtccgat cgtgacggcg atcctcgatg cggacacacc gttcgacgag ttccaggcca
34081 attgggcgcc gtcgatggtg gtcgggtgg gtcggctgc gggtcgcacg gtgggtgtac
34141 tggccaacaa cccgctacgc ctggcgccgt gcctgaactc cagaagcgca gagaaggcag
34201 cgcgttcgtg cggcgctgtc gacgcgttcg ggattccgt ggtgggtgt gtcgatgtc
34261 cgggctatct gcccgggtgc gaccaggagt ggggtggcgt ggtgcgcgt ggcgccaagt
15 34321 tgcgtcacgc gtccggcgag tgcacgcgtc cgccgggtcac gtcggtcacc cgaaagaccc
34381 acggcgcccc atacattgcg atgaactccc gtcgttgaa cgcgaccaag gtgtcgcc
34441 ggcggacgc cgaggtcgcg gtgatggcg ctaaggcgac ctcgcgtac tcgcacaaga
34501 agaagttggc cgccgctccg gaggcacgaac gcaagcgct gcacgaccag ttggccgccc
34561 agcatgagcg catcgccggc ggggtcgaca gtgcgttgaa catcggtgt gtcgacgaga
20 34621 agatcgaccc ggcgcatact cgcagcaac tcaccgaggc gtcggcgac gtcggccac
34681 ggcgcggccg ccacaagaac atcccgtgt agttctgacc gcgagcagac gcagaatcgc
34741 acgcgcgagg tccgcgcgt gcgattctgc gtcgtcgcc cagttatccc cagcggtggc
34801 tggtaacgc gaggcgctcc tcgcgtgtc ggacgggtgcc taccgacgcg ctaacaattc
34861 tcgagaaggc cggcggttc gccaccaccg cgcaattgtc cacggtcatg acccgccaac
25 34921 agtcgcacgt ccaagtggaa aacggcgcc tcgttcgt tgggtacggg gtctacgcgg
34981 cacaagagcc ggacctgttg ggccgcgttgg cggctctcgat tgggtcatg ggggggcacg
35041 ccgtcgctgt tctggccacc gccgcgcgt tggatggatt cgacacggaa aacaccgtcg
35101 ctatccatat gtcgtatccc ggagtaaggc tgccggccac ggtcggtctg atggccacc
35161 aacgcgtcg tgccggctc caacgggtgt caggtcgat cgcgaccgcg cccgcatgg
30 35221 ctgcgttggc ggtcgacga cagttcgcc gcccggggc gtcggccacc ctcgcacgcg
35281 cactacggtc aatgcgtgc gtcgcgtg aaattggaa cgcgcgttgc gagcagcgc
35341 gccgcgcagg catcgctcg ggcgcgaac tcttaccctt cggccacggc cgcgcggaaat
35401 cggccatggc gagcgaggct cggctcgat tgatcgacca cgggtgcgg ttggccgaac
35461 ttcaataccat gatacacggc cacgggtgtt aaatgtggcg agtgcacttc gcctggcccg
35521 acatgcgtct cgcggccgaa tacgaaagca tcgagtgcc cgcgggaccc gcgagatgc
35581 tgcgcgacaa gacacgctgg gccaaatcc aagagctcg gttggacgattt gtcggattt

35641 tcgtcgacga tgcagacgc gaacccggcc gcctggcggc ccgcacgcgc cgccacctcg
35701 accgcgcgac tatggccggc tgaccgtgg tgagcagacg cagactcgca ctgcggccgg
35761 cgcagtgcga ctctgcgtct gtcgcgtc aacggctgag gaactcccta gccacggcga
5 35821 ctacgcgtc gcgatcccgt ggcaccagac cgatccgggt ccggcggtcg aggatatcg
35881 ccacatccag cgcggccctca tgggtcacgg cgtattcgaat ctccggccgg gtacacgtcga
35941 tgcgtcgcc gaccggctcg gtggccgct cacatgtggc ggcggcagcg acgtggccg
36001 cctcgcccccc gtaccgcgc accagcgact cgggaatcc ggcggccgat cggggggccg
36061 gcccagggtt cgccgggtcgcc cgcgtcgcgc gcagggttgcg agtgcggcac ttgcggcgtc
10 36121 gcagggtcgc cagcgtgatg gcgcgattca gcacatccctc tgccatgttag cggtattccg
36181 tcagttgcc gccgaccaca ctgatcacgc cgcacggcga ttcaaaaaca gcgtggtcac
36241 gcgaaacgtc ggcgggtcgcc ccctggacac cagcaccggc ggtgtcgatt agcggccgca
36301 atcccgata ggcaccggatg acatccctgg tgccgaccgc cgtcccaat gcgggttca
36361 ccgtatcccg caggaacgtg atctttccg aagacgggttgcg acgtcgcgc ggaatcggc
15 36421 cgggtgcgtc ttgcgtcgcc agcccgagat agatccggcc cagctgcgtc ggcacatggcga
36481 acacgaagcg gttcagctca cggggatcg gaatggtcag cgcggcagtc ggattggcaa
36541 acgacttcgc gtcgaagacc agatgtgtgc cgcggctggg gcgtgcgcctc agggacgggt
36601 ccatctcacc cgcacacacg cccgcccgt tgatgacggc acgcgcgcac agcgcgaacg
36661 actgcccgggt ggcgggtcg gtcaactcca ccgaagtgcgc ggtgacattc gacgcgcaca
20 36721 cgtaaagttag gatgcggggcg ccgtgtggg ccgcgggtcg cgcacggcc atgaccagcc
36781 gggcgtcgatc gatcaattgc ccgtcgatcg cgagcagacc accgtcgagg ccgtcccgcc
36841 gaacgggtggg agcaatctcc accacccgtg acgcggggat tcggcgcgtat cggggcaacg
36901 tcgcccggc cgtacccgtc agcaccgc aagcgtcgcc ggccaggaaa cggcacgc
36961 ccaacgcggc cttgggtgtga cccatcgacg gcaacaacgg gaccagtgc ggcacatggcat
25 37021 gcacgagatg aggagcggttg cgtgtcatca ggattccgcg ttgcacggcg ctgcggccgg
37081 ccatgcccac gttggccgtc gccagatgc gcagaccgcgt gtgcaccaac ttgcagctcc
37141 agcggctggc ggcgaacgc agatcatgttttccacaa ggcacccgtc agaccgcggg
37201 tggcagcatc taaggcaatg ccaacacccggtaatgcgcgc gctatcactg atgacgtcga
37261 gtgcggccacc gtcggccact gcggtcaggat cggcggagcgc acgcgcgcgc ttgagtgcag
30 37321 ccgagtggttgc catcagcaca aatatccgtt cagtcgtgg gtaagttcg tgccagcgc
37381 ggcggaaatcg aggatcgaat cgacgtgc cgcggactgg atggtcact gggcgtcag
37441 caacaccatg gtgcggcgtc gacgagcgtc gccggagcgc acactgcggc accgcgtcgc
37501 cactgtcagc cggggggccca acccctcgat caggacctgc tggtggcgtc cgaggcgc
37561 ggtgtatgc accctggcca gtcggcgttgc catgaccgc acgtcaccatc
37621 caaccggcgtc gccaccgcga caaatgtgtt taccaacgc tcccggtcgt ccccggtcag
35 37681 gggcacccatc cgcagcactg cggcgtatgc gtcggcgtc atggacgcgcata tgatcgaccg

37741 ggtgtccggc cagcgacggt atacggtcgg ggggctacg cccgcgcgcc gggcgatct
37801 ggcaagtgtc accccggtcca cgccgttaatc gacgacgcag ctcgcccgtg cccgcaggat
37861 acgaccacccg gtatccgcgc ggtcattact caattgacagc atgtgtaaata ctgtaacgcg
37921 tgacttcacccg cgaggaactc cttccaccga tgaaaatggga cgcgtgggga gatccgcgg
37981 cggccaagcc actttctgtat ggctgtccggc cgttgctgaa gcagggtgtg ggcctagcgg
38041 actcggagca gcccgaactc gacccgcgc aggtgcagct ggcgcgtcc gccctgtcgg
38101 gggcagacca (SEQ ID NO: 24)

6.9. X-linked Inhibitor of Apoptosis Protein ("XIAP")

GenBank Accession # U45880:

1 gaaaagggtgg acaagtccta tttcaagag aagatgacit ttaacagttt tgaaggatct
61 aaaacttgt tacctgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga
121 taaaaaactt ttgctaattt tccaagtgg agtcctgtt cagcatcaac actggcagca
181 gcagggtttc ttatactgg tgaaggagat accgtgcggt gcttagttg tcatgcagct
241 gtagatagat ggcaaatatgg agactcagca gtggaaagac acagggaaagt atccccaaat
301 tgcagattt tcaacggctt ttatcttggaa aatagtgccca cgcgactctac aaattctggt
361 atccagaatg gtcagtacaa agttgaaaac tatctggaa gcagagatca tttgcctt
421 gacaggccat ctgagacaca tgcagactat ctttgagaa ctggcaggt tgtagatata
481 tcagacacca tatacccgag gaaccctgcc atgtatttgt aagaagctag attaaagtcc
541 tttcagaact ggccagacta tgctcacca accccaagag agtttagcaag tgctggactc
601 tactacacag gtatttgta ccaagtgcag tgcttttgtt gtgggtggaaa actgaaaaat
661 tgggaacctt gtatcgtgc ctggtcagaa cacaggcgcac actttcttaa ttgtcttctt
721 gtttggcc ggaatctaa tattcgaagt gaatctgatg ctgtgagttc tgataggaat
781 ttcccaaattt caacaatct tccaagaaat ccatccatgg cagattatga agcacggatc
841 ttacttttggacatggat atactcgat aacaaggagc agcttgcag agctggattt
901 tatgcttttag gtgaaggtga taaagttaaag tgcttcaact gtggaggagg gctaactgtat
961 tggaaagccca gtgaagaccc ttggaaacaa catgtaaat ggtatccagg gtgcacaaat
1021 ctgttagaac agaagggaca agaataatata aacaatatttc atttaactca ttcacitgag
1081 gagtgtctgg taagaactac tgagaaaaca ccatcaactaa cttagaagaat tgatgatacc
1141 atcttccaaa atccatggt acaagaagct atacgaatgg gggtcagttt caaggacatt
1201 aagaaaataa tggaggaaaa aattcagata tctggggagca actataaattc acttgagggtt
1261 ctggtgtcag atctgtgaa tgctcagaaa gacagtatgc aagatgagtc aagtcagact
1321 tcattacaga aagagattag tactgaagag cagctaaggc gcctgcaaga ggagaagctt
1381 tgccaaatct gtatggatag aaatattgtc atcggttttgc ttccctgtgg acatctgatc
1441 acttgtaaac aatgtgctga agcagttgac aagtgtccca tgcgtctac acgtcattact

1501 ttcaagcaaa aaattttat gtcttaatct aactctata taggcattttt atgttgttct
 1561 tattaccctg aitgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat
 1621 tagcatttgc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata
 5 1681 atcttgaaat ttcttgattt tcagggtat tagctgtatt atccattttt ttactgtta
 1741 tttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt
 1801 attcatagta tactgatttta atttctaagt gtaagtgaat taatcatctg gatttttat
 1861 tctttcaga taggcttaac aaatggagct ttctgtatataaaatgtggag attagagtt
 1921 atctccccaa tcacataattt tgtttgggtt gaaaaaggaa taaattgttc catgctgggt
 1981 gaaagataga gattgtttt agagggtgg tttttttt taggattctg tccattttct
 10 2041 tgtaaaggga taaacacgga cgtgtgcgaa atatgtttgtt aaagtgtttt gccatttttgg
 2101 aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgtca
 2161 gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatgtctg agccagatca
 2221 aagtatgtat ttgttaata tgcataaaac gagagatgg gaaagatata caccaaatcg
 2281 ttaaatgtgg ttctcttcg gggagggggg gattttttttt gggggccccc agggggttttt
 2341 gagggccctt ttacttttcg actttttca ttttttctg ttccggattttt ttataagtat
 2401 gtagaccccg aagggtttt tgggaactaa catcagtaac ctaaccccg tgactatcc
 2461 gtgctttcc tagggagctg ttgtttcc caccaccac cttccctct gaacaaatgc
 2521 ctgagtgtcg gggcactttg (SEQ ID NO: 25)

20 25 General Target Region:

Internal Ribosome Entry Site (IRES) in 5' untranslated region:

5'AGCUCCUAUAACAAAAGUCUGUUGCUUGUGUUUCACAUUUUGGAUU
 UCCUAAUAAAUGUUCUCUUUUUAGAAAAGGUGGGACAAGUCCUAUUU
 UCAAGAGAAG3' (SEQ ID NO: 26)

25

Initial Specific Target Motif:

RNP core binding site within XIAP IRES

5'GGAUUUCUAAUAAAUGUUCUCUUUUU3' (SEQ ID NO: 27)

30

6.10. Survivin

GenBank Accession # NM_001168:

1 ccggccagatt tgaatcgccg gaccgggtgg cagaggtggc ggccggccca tgggtgcccc
 61 gacgttgcggc cctgcctggc agcccttct caaggaccac cgcatctta cattcaagaa
 35 121 ctggcccttc ttggagggtt ggcgcctgcac cccggagccg atggccgagg ctggcttcat
 181 ccactgcccc actgagaacg agccagactt ggcccgatgt ttcttctgct tcaaggagct

241 ggaaggctgg gagccagatg acgaccat agaggaacat aaaaagcatt cgtccggttg
 301 cgcttcctt tctgtcaaga agcagttga agaattaacc ttggtaat tttgaaact
 361 ggacagagaa agagccaaga acaaattgc aaaggaaacc aacaataaga agaaagaatt
 5 421 tgaggaaact gcbaagaaag tgcgccgtgc catcgagcag ctggctgcca tggattgagg
 481 cctctggccg gagctgcctg gtcccagagt ggctgcacca cttccagggt ttattccctg
 541 gtgccaccag cttccctgtg ggcccccttag caatgtctta ggaaggaga tcaacattt
 601 caaatttagat gttcaactg tgctccgtt ttgtctgaa agtggcacca gaggtgtc
 661 tgcctgtgca gcgggtgctg ctggtaacag tggctgcttc tctctctc tctctttt
 721 ggggctcat tttgctgtt tgattcccg gccttaccag gtgagaagtg agggaggaag
 10 781 aaggcagtgt ccctttgtc agagctgaca gcttigtcg cgtggcaga gcctccaca
 841 gtgaatgtgt ctggacctca tggatgttgg gctgtcacag tcctgatgtt ggacttgca
 901 ggtgcctgtt gaatctgagc tgcagggtcc ttatctgtca cacctgtgcc tcctcagagg
 961 acagttttt tggatgttgg tttttttttt gtagatgca tgacttgtt
 1021 gtgatgagag aatggagaca gagtccctgg ctccctact gtttacaac atggtttct
 1081 tattttgtt gaatgttaa ttcacagaat agcacaact acaattaaaa ctaagcaca
 1141 agccattcta agtcattggg gaaacggggt gaacttcagg tggatgagga gacagaatag
 1201 agtgatagga agcgtctggc agatactcct ttgcactg ctgtgtgatt agacaggccc
 1261 agtgagccgc ggggcacatg ctggccgtc ctccctcaga aaaaggcagt ggcctaaatc
 20 1321 cttttaaat gacttggctc gatgctgtgg gggactggct gggctgtc aggccgtgt
 1381 tctgtcagcc caacccac atctgtcacg ttctccacac gggggagaga cgcagtccgc
 1441 ccagggtcccc gtttcttgg gaggcagcag ctcccgagg gctgaagtct gggttaagat
 1501 gatggatttgcattcgccctc ctccctgtca tagagctgca gggtgatttgcatttc
 1561 gctggaaacc tctggaggc atctcggtc ttccctgagaa ataaaaagcc tgcatttc (SEQ ID NO: 28)

25

**7. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA
BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS**

The results presented in this Example indicate that gel mobility shift assays can be used to detect the binding of small molecules, such as the Tat peptide and gentamicin, to their respective target RNAs.

7.1. Materials and Methods

7.1.1. Buffers

35 Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl₂. Tris-borate-EDTA (TBE) buffer is

composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl₂.

5

7.1.1. Gel retardation analysis

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of either a 5' fluorescein labeled oligonucleotide corresponding to the 16S rRNA A site (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 29); Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448) or a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was 3' labeled with 5'-³²P cytidine 3', 5'-bis(phosphate) (NEN) and T4 RNA ligase (NEBiolabs) in 10% DMSO as per manufacturer's instructions. The labeled oligonucleotides were purified using G-25 Sephadex columns (Boehringer Mannheim). For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was utilized with TK buffer containing 0.5mM MgCl₂ and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 31); single letter amino acid code). For 16S rRNA-gentamicin reactions, the method of Huq *et al.* was used with TKM buffer. In 20 µl reaction volumes 50 pmoles of ³²P cytidine-labeled oligonucleotide and either gentamicin sulfate (Sigma) or the short Tat peptide (Tat₄₇₋₅₈) in TK or TKM buffer were heated at 90°C for 2 minutes and allow to cool to room temperature (approximately 24°C) over 2 hours. Then 10 µl of 30% glycerol was added to each reaction tube and the entire sample was loaded onto a TBE non-denaturing polyacrylamide gel and electrophoresed at 1200-1600 volt-hours at 4°C. The gel was exposed to an intensifying screen and radioactivity was quantitated using a Typhoon phosphorimager (Molecular Dynamics).

30

7.2. Background

One method used to demonstrate small molecule interactions with natural occurring RNA structures such as ribosomes is by a method called chemical footprinting or toe printing (Moazed & Noller, 1987, Nature 327:389-394; Woodcock *et al.*, 1991, EMBO J. 10:3099-3103; Yoshizawa *et al.*, 1998, EMBO J. 17:6437-6448). Here the use of gel mobility shift assays to monitor RNA-small molecule interactions are described. This approach allows for rapid visualization of small molecule-RNA interactions based on the

difference between mobility of RNA alone versus RNA in a complex with a small molecule. To validate this approach, an RNA oligonucleotide corresponding to the well-characterized gentamicin binding site on the 16S rRNA (Moazed & Noller, 1987, *Nature* 327:389-394) and the equally well-characterized HIV-1 TAT protein binding site on the HIV-1 TAR element (Huq *et al.*, 1999, *Nucleic Acids Res.* 27: 1084-1093) were chosen. The purpose of these experiments is to lay the groundwork for the use of chromatographic techniques in a high throughput fashion, such as microcapillary electrophoresis, for drug discovery.

10

7.3. Results

A gel retardation assay was performed using the Tat₄₇₋₅₈ peptide and the TAR RNA oligonucleotide. As shown in Figure 1, in the presence of the Tat peptide, a clear shift is visible when the products are separated on a 12% non-denaturing polyacrylamide gel. In the reaction that lacks peptide, only the free RNA is visible. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, *Proc. Natl. Acad. Sci. USA* 94:3548-3553; Huq *et al.*, 1999, *Nucleic Acids Res.* 27: 1084-1093).

Based on the results of Figure 1, it was hypothesized that RNA interactions with small organic molecules could also be visualized using this method. As shown in Figure 2, the addition of varying concentrations of gentamicin to an RNA oligonucleotide corresponding to the 16S rRNA A site produces a mobility shift. These results demonstrate that the binding of the small molecule gentamicin to an RNA oligonucleotide having a defined structure in solution can be monitored using this approach. In addition, as shown in Figure 2, a concentration as low as 10ng/ml gentamicin produces the mobility shift.

To determine whether lower concentrations of gentamicin would be sufficient to produce a gel shift, similar experiment was performed, as shown in Figure 2, except that the concentrations of gentamicin ranged from 100 ng/ml to 10 pg/ml. As shown in Figure 3, gel mobility shifts are produced when the gentamicin concentration is as low as 10 pg/ml. Further, the results shown in Figure 3 demonstrate that the shift is specific to the 16S rRNA oligonucleotide as the use of an unrelated oligonucleotide, corresponding to the HIV TAR RNA element, does not result in a gel mobility shift when incubated with 10 µg/ml gentamicin. In addition, if a concentration as low as 10 pg/ml gentamicin produces a gel mobility shift then it should be possible to detect changes to RNA structural motifs when small amounts of compound from a library of diverse compounds is screened in this fashion.

Further analysis of the gentamicin-RNA interaction indicates that the interaction is Mg²⁺- and temperature dependent. As shown in Figure 4, when MgCl₂ is not present (TK buffer), 1mg/ml of gentamicin must be added to the reaction to produce a gel shift.

Similarly, the temperature of the reaction when gentamicin is added is also important. When gentamicin is present in the reaction during the entire denaturation/renaturation cycle, that is, when gentamicin is added at 90°C or 85°C, a gel shift is visualized (data not shown). In contrast, when gentamicin is added after the renaturation step has proceeded to 75°C, a mobility shift is not produced. These results are consistent with the notion that gentamicin may recognize and interact with an RNA structure formed early in the renaturation process.

**8. EXAMPLE: IDENTIFICATION OF A DYE-LABELED TARGET RNA
15 BOUND TO SMALL MOLECULAR WEIGHT COMPOUNDS
BY CAPILLARY ELECTROPHORESIS**

The results presented in this Example indicate that interactions between a peptide and its target RNA, such as the Tat peptide and TAR RNA, can be monitored by gel retardation assays in an automated capillary electrophoresis system.

20 8.1. Materials and Methods

8.1.1. Buffers

Tris-potassium chloride (TK) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100, and 0.5mM MgCl₂. Tris-borate-EDTA (TBE) buffer is composed of 45 mM Tris-borate pH 8.0, and 1 mM EDTA. Tris-Potassium chloride-magnesium (TKM) buffer is composed of 50 mM Tris-HCl pH 7.4, 20mM KCl, 0.1%Triton X-100 and 5mM MgCl₂.

30 8.1.1. Gel Retardation Analysis Using Capillary Electrophoresis

RNA oligonucleotides were purchased from Dharmacon, Inc, Lafayette, CO). 500 pmole of a 5' fluorescein labeled oligonucleotide corresponding to the HIV-1 TAR element TAR RNA (5'-GGCGUCACACCUCUUCGGGUGAAGUCGCC-3' (SEQ ID NO: 30); Huq *et al.*, 1999, Nucleic Acids Research. 27:1084-1093; Hwang *et al.*, 1999, Proc. Natl. Acad. Sci. USA 96:12997-13002) was used. For Tat-TAR gel retardation reactions the method of Huq *et al.* (Nucleic Acids Research, 1999, 27:1084-1093) was

utilized with TK buffer containing 0.5mM MgCl₂ and a 12-mer Tat peptide (YGRKKRRQRRRP (SEQ ID NO: 31); single letter amino acid code). In 20 μ l reaction volumes 50 pmoles of labeled oligonucleotide and the short Tat peptide (Tat₄₇₋₅₈) in TK or 5 TKM buffer were heated at 90 °C for 2 minutes and allow to cool to room temperature (approximately 24 °C) over 2 hours. The reactions were loaded onto a SCE9610 automated capillary electrophoresis apparatus (SpectruMedix; State College, Pennsylvania).

8.2. Results

As presented in the previous Example in Section 7, interactions between a peptide and RNA can be monitored by gel retardation assays. It was hypothesized that interactions between a peptide and RNA could be monitored by gel retardation assays by an automated capillary electrophoresis system. To test this hypothesis, a gel retardation assay by an automated capillary electrophoresis system was performed using the Tat₄₇₋₅₈ peptide 10 and the TAR RNA oligonucleotide. As shown in Figure 5 using the capillary electrophoresis system, in the presence of the Tat peptide, a clear shift is visible upon the addition of increasing concentrations of Tat peptide. In the reaction that lacks peptide, only 15 a peak corresponding to the free RNA is observed. These observations confirm previous reports made using other Tat peptides (Hamy *et al.*, 1997, Proc. Natl. Acad. Sci. USA 94:3548-3553; Huq *et al.*, 1999, Nucleic Acids Res. 27: 1084-1093). 20

The present invention is not to be limited in scope by the specific 25 embodiments described herein. Indeed, various modifications of the invention in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties.

30

35

The invention can be illustrated by the following embodiments enumerated in the numbered paragraphs that follow:

- 5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed; (b) separating the detectably labeled target RNA:test compound complex formed in step(a) from uncomplexed target RNA molecules and test compounds; and (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex.
- 10 2. The method of paragraph 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or adenylate uridylate-rich element.
- 15 3. The method of paragraph 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.
- 20 4. The method of paragraph 1 in which the detectably labeled RNA is labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.
- 25 5. The method of paragraph 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries, including but not limited to, libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

6. The method of paragraph 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution, the aqueous solution comprising a buffer and a combination of salts, preferably approximating or mimicking physiologic conditions.

5

7. The method of paragraph 6 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

10

8. The method of paragraph 6 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant. In another embodiment, the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂. In a preferred embodiment, the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂. In another embodiment, the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

15

9. Any method that detects an altered physical property of a target nucleic acid complexed to a test compound from the unbound target nucleic acid may be used for separation of the complexed and non-complexed target nucleic acids in the method of paragraph 1. In a preferred embodiment, electrophoresis is used for separation of the complexed and non-complexed target nucleic acids. In a preferred embodiment, the electrophoresis is capillary electrophoresis. In other embodiments, fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, and nanoparticle aggregation are used for the separation of the complexed and non-complexed target nucleic acids.

20

25

10. The structure of the test compound of the RNA:test compound complex of paragraph 1 is determined, in part, by the type of library of test compounds. In a preferred embodiment wherein the combinatorial libraries are small organic molecule libraries, mass spectroscopy, NMR, or vibration spectroscopy are used to determine the structure of the test compounds.

30

35

WHAT IS CLAIMED IS:

1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:
 - 5 (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed;
 - 10 (b) separating the detectably labeled target RNA:test compound complex formed in step(a) from uncomplexed target RNA molecules and test compounds by capillary gel electrophoresis; and
 - 15 (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex by mass spectroscopy.

20

25

30

35

AMENDED CLAIMS

[received by the International Bureau on 17 September 2002 (17.09.02);
Claims 1 to 10 replaced by new claims 1 to 19. (3 sheets)]

5 1. A method for identifying a test compound that binds to a target RNA molecule, comprising the steps of:

- (a) contacting a detectably labeled target RNA molecule with a library of test compounds under conditions that permit direct binding of the labeled target RNA to a member of the library of test compounds so that a detectably labeled target RNA:test compound complex is formed;
- (b) separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA molecules and test compounds; and
- 15 (c) determining a structure of the test compound bound to the RNA in the RNA:test compound complex.

2. The method of claim 1 in which the target RNA molecule contains an HIV TAR element, internal ribosome entry site, "slippery site", instability element, or
20 adenylate uridylate-rich element.

3. The method of claim 1 in which the RNA molecule is an element derived from the mRNA for tumor necrosis factor alpha ("TNF- α "), granulocyte-macrophage colony stimulating factor ("GM-CSF"), interleukin 2 ("IL-2"), interleukin 6
25 ("IL-6"), vascular endothelial growth factor ("VEGF"), human immunodeficiency virus I ("HIV-1"), hepatitis C virus ("HCV" - genotypes 1a & 1b), ribonuclease P RNA ("RNaseP"), X-linked inhibitor of apoptosis protein ("XIAP"), or survivin.

4. The method of claim 1 in which the detectably labeled RNA is
30 labeled with a fluorescent dye, phosphorescent dye, ultraviolet dye, infrared dye, visible dye, radiolabel, enzyme, spectroscopic colorimetric label, affinity tag, or nanoparticle.

5. The method of claim 1 in which the test compound is selected from a combinatorial library comprising peptoids; random bio-oligomers; diversomers such as
35 hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal

peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; or small organic molecule libraries.

5

6. The method of claim 5 in which the small organic molecule libraries are libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

10

7. The method of claim 1 in which screening a library of test compounds comprises contacting the test compound with the target nucleic acid in the presence of an aqueous solution wherein the aqueous solution comprises a buffer and a combination of salts.

15

8. The method of claim 7 wherein the aqueous solution approximates or mimics physiologic conditions.

9. The method of claim 7 in which the aqueous solution optionally further comprises non-specific nucleic acids comprising DNA, yeast tRNA, salmon sperm DNA, homoribopolymers, and nonspecific RNAs.

10. The method of claim 7 in which the aqueous solution further comprises a buffer, a combination of salts, and optionally, a detergent or a surfactant.

25

11. The method of claim 10 in which the aqueous solution further comprises a combination of salts, from about 0 mM to about 100 mM KCl, from about 0 mM to about 1 M NaCl, and from about 0 mM to about 200 mM MgCl₂.

30

13. The method of claim 11 wherein the combination of salts is about 100 mM KCl, 500 mM NaCl, and 10 mM MgCl₂.

14. The method of claim 10 wherein the solution optionally comprises from about 0.01% to about 0.5% (w/v) of a detergent or a surfactant.

35

15. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and
5 test compounds is by electrophoresis.

16. The method of claim 15 in which the electrophoresis is capillary electrophoresis.

10 17. The method of claim 1 in which separating the detectably labeled target RNA:test compound complex formed in step (a) from uncomplexed target RNA and test compounds is by fluorescence spectroscopy, surface plasmon resonance, mass spectrometry, scintillation, proximity assay, structure-activity relationships ("SAR") by NMR spectroscopy, size exclusion chromatography, affinity chromatography, or
15 nanoparticle aggregation.

18. The method of claim 1 in which the library of test compounds are small organic molecule libraries.

20 19. The method of claim 18 in which the structure of the test compound is determined by mass spectroscopy, NMR, or vibration spectroscopy.

Figure 1
Sheet 1/5
Attorney Docket No. 10589-007

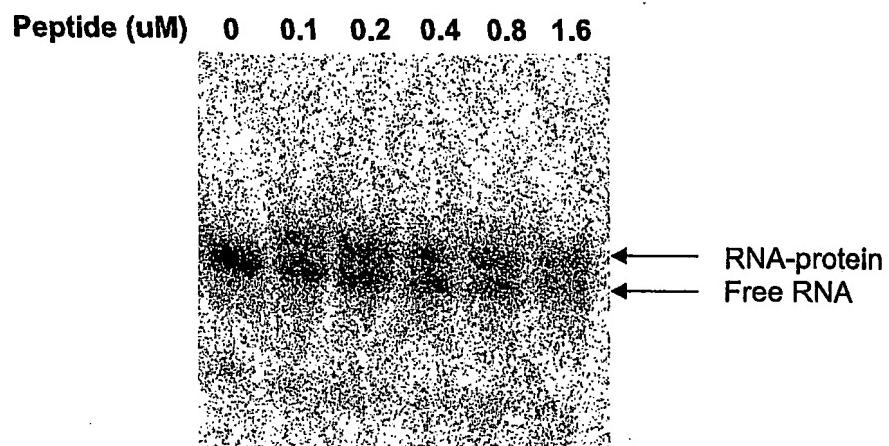


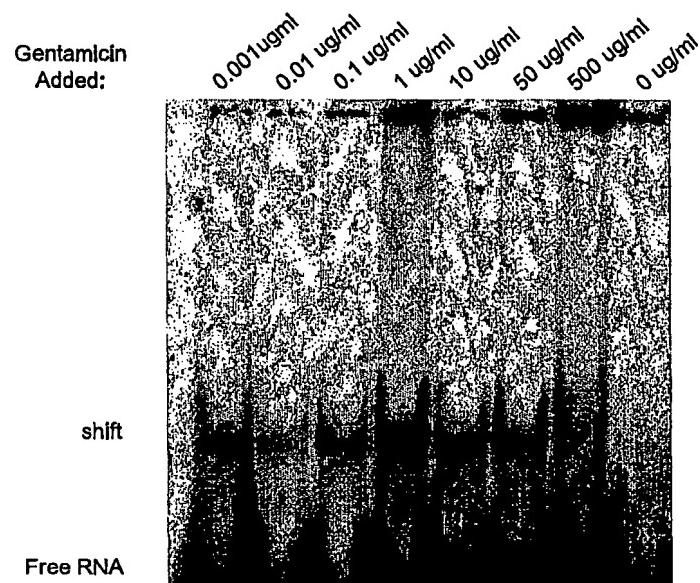
Figure 2**Sheet 2/5****Attorney Docket No. 10589-007**

Figure 3
Sheet 3/5
Attorney Docket No. 10589-007

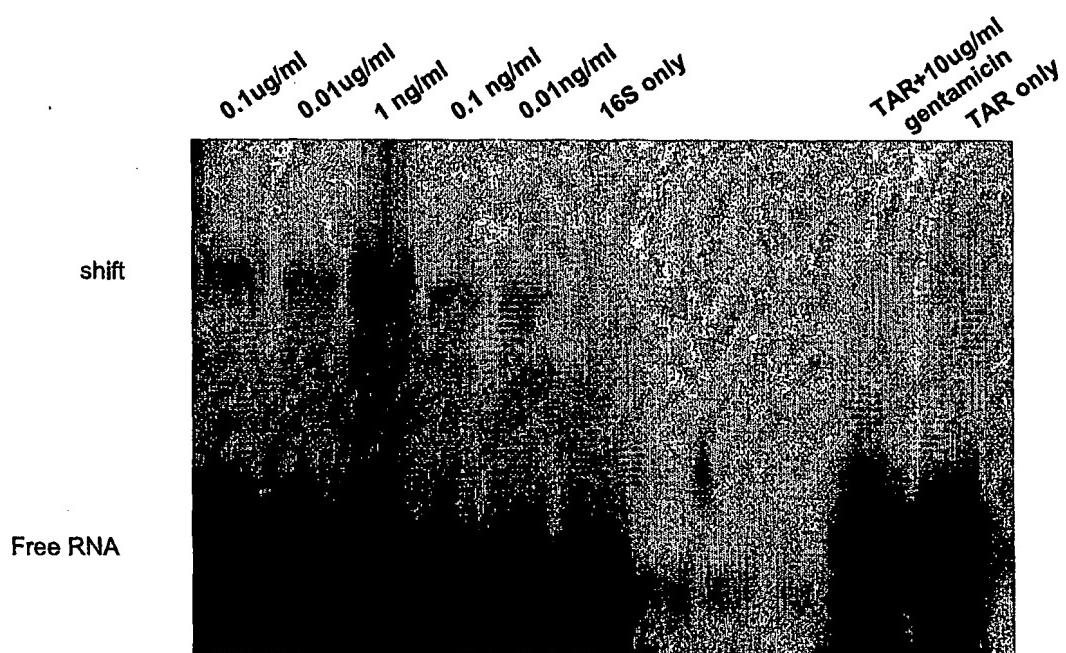


Figure 4
Sheet 4/5
Attorney Docket No. 10589-007

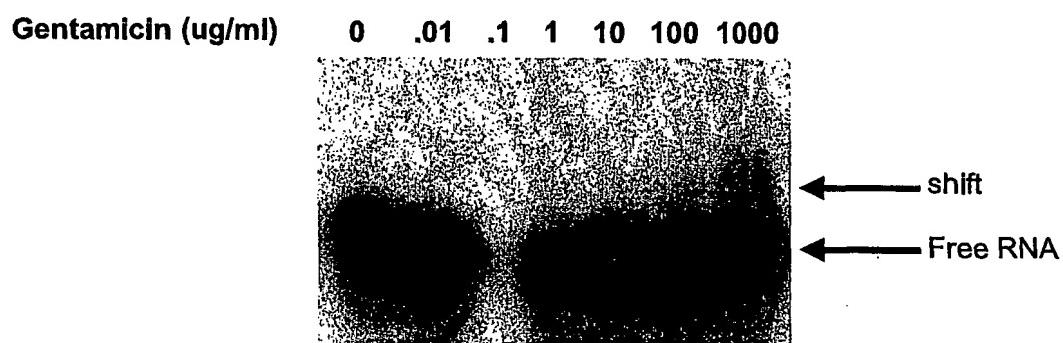
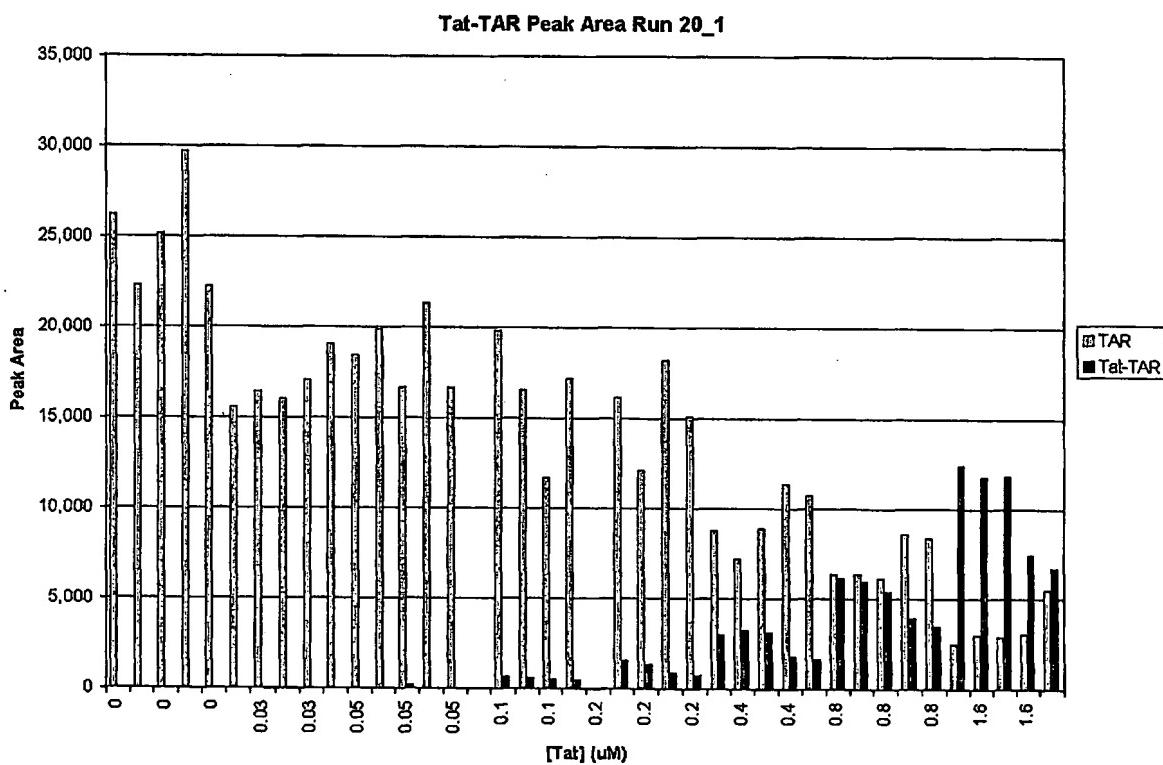


Figure 5
Sheet 5/5
Attorney Docket No. 10589-007



SEQUENCE LISTING

<110> PTC Therapeutics, Inc.

<120> METHODS FOR IDENTIFYING SMALL MOLECULES THAT BIND SPECIFIC RNA STRUCTURAL MOTIFS

<130> 10589-007-228

<140> To be assigned

<141> 2002-04-11

<150> 60/282,965

<151> 2001-04-11

<160> 31

<170> PatentIn version 3.0

<210> 1

<211> 21

<212> RNA

<213> Homo sapiens

<400> 1

auuuauuuau uuauuuauuu a

21

<210> 2

<211> 17

<212> RNA

<213> Homo sapiens

<400> 2

auuuauuuau uuauuuua

17

<210> 3
<211> 15
<212> RNA
<213> Homo sapiens

<400> 3
wauuuuauuuua uuuuaw

15

<210> 4
<211> 13
<212> RNA
<213> Homo sapiens

<400> 4
wwauuuauuu aww

13

<210> 5
<211> 13
<212> RNA
<213> Homo sapiens

<400> 5
wwwwaauuaw www

13

<210> 6
<211> 1643
<212> DNA
<213> Homo sapiens

<400> 6
gcagaggacc agctaagagg gagagaagca actacagacc cccccctgaaa acaaccctca 60
gacgccacat cccctgacaa gctgccaggc aggttctctt cctctcacat actgaccac 120
ggctccaccc tctctcccct ggaaaggaca ccatgagcac tgaaagcatg atccgggacg 180
tggagctggc cgaggaggcg ctccccaga agacaggggg gccccagggc tccaggcggt 240
gcttgccct cagccctttc tccttcctga tcgtggcagg cgccaccacg ctcttctgcc 300

tgctgcactt tggagtgatc ggccccaga gggaaagagtt ccccaggac ctctctctaa	360
tcagccctct ggcccaggca gtcagatcat ctctcgAAC cccgagtgtac aagcctgtAG	420
cccatgttgt agcaaaccct caagctgagg ggcagctcca gtggctgaAC cgccgggcca	480
atgccctcct ggccaatggc gtggagctga gagataacca gctggtggtG ccatcagagg	540
gcctgtacct catctactcc caggtcctct tcaaggGCCA aggctcccccc tccacccatG	600
tgctcctcac ccacaccatc agccgcATCG ccgtctccta ccagaccaAG gtcaacctcc	660
tctctgcccAt caagagcccc tgccagaggg agacccaga gggggctgag gccaagccct	720
ggtagatGCC catctatctG ggaggggtct tccagctgGA gaagggtgac cgactcagcg	780
cTGAGATCAA TCGGCCGAC tatctcgact ttGCCGAGTC tGGCAGGTC tactttggGA	840
tcattggcct gtgaggagga cgaacatCCA accttccAA acgcctcccc tgccccaaatc	900
cctttattac cccctccttc agacaccctc aacctttct ggctaaaaa gagaattggg	960
ggcttagggt cggaacccaa gcttagaact ttaagcaaca agaccaccac ttGAAACCT	1020
gggattcagg aatgtgtggc ctgcacagtG aattgctggc aaccactaAG aattcaaACT	1080
ggggcctCCA gaactcactg gggcctacAG ctGatccc tgacatctgg aatctggaga	1140
ccaggagCC tttggttctG gCcAGAATGC tGcAGGACTT gagaAGACCT cacCTAGAAA	1200
ttGACACAAG tggacacTTAG gccttcctct ctccAGATGT ttccAGACTT ccttgagaca	1260
cgagccccAG ccctccccat ggagccAGt ccctctatTT atGTTGcAc ttGtGATTAT	1320
ttattattta ttattatttt atttattttAC agatGAATGT atttattttgg gagaccgggg	1380
tatccTgggg gaccaatgt aggAGCTGCC ttggctcaga catGTTTCC gtGAAAACGG	1440
agctGAACAA taggCTGTTc ccatGTAgeCC ccctggcCtc tGcCTTCT tttgattatG	1500
ttttttaaaa tatttatctG attaAGTTGT ctaaacaatG ctGATTGgt gaccaactgt	1560
cactoattgc tgAGCCTCTG ctccccAGGG gagttgtgtc tgtaatcgCC ctactattca	1620
gtggcgagaa ataaAGTTG CTT	1643

<210> 7

<211> 756

<212> DNA

<213> Homo sapiens

<400> 7	
gctggaggat gtggctgcAG agcctgctgc tcttggcAC tGtggcctgc agcatctctG	60
cacccggcccg ctGcccAGc cccAGCACGc agccctggGA gcatgtGAAT gccatCCAGG	120
aggccccggcg tctcctgaAC ctgagtagAG acactgctgc tgagatGAAT gaaacAGTAG	180

aagtcatctc agaaatgtt gacccctcagg agccgacctg cctacagacc cgccctggagg	240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg	300
ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtgca acccagacta	360
tcacccttga aagtttcaaa gagaacctga aggactttct gcttgcata cccttgact	420
gctggagcc agtccaggag tgagaccggc cagatgaggo tggccaagcc ggggagctgc	480
tctctcatga aacaagagct agaaactcag gatggcata ttggaggac caaggggtgg	540
gccacagcca tggggaggt ggcctggacc tgccctgggc cacactgacc ctgatacagg	600
catggcagaa gaatggaaat attttatact gacagaaatc agtaatattt atatatttat	660
atttttaaaa tatttattta ttatatttata taagttcata ttccatattt attcaagatg	720
ttttaccgta ataattatta taaaaatat gcttct	756

<210> 8

<211> 756

<212> DNA

<213> Homo sapiens

<400> 8	
tctggaggat gtggctgcag agcctgctgc tcttggcac tgtggcctgc agcatctctg	60
cacccgccccg ctcgcccagc cccagcacgc agccctggga gcatgtaat gcoatccagg	120
aggcccccgcg tctcctgaac ctgagtagag acactgctgc tgagatgaat gaaacagtag	180
aagtcatctc agaaatgtt gacccctcagg agccgacctg cctacagacc cgccctggagg	240
tgtacaagca gggcctgcgg ggcagcctca ccaagctcaa gggccccttg accatgatgg	300
ccagccacta caagcagcac tgccctccaa ccccgaaac ttccctgtgca acccagacta	360
tcacccttga aagtttcaaa gagaacctga aggactttct gcttgcata cccttgact	420
gctggagcc agtccaggag tgagaccggc cagatgaggo tggccaagcc ggggagctgc	480
tctctcatga aacaagagct agaaactcag gatggcata ttggaggac caaggggtgg	540
gccacagcca tggggaggt ggcctggacc tgccctgggc cacactgacc ctgatacagg	600
catggcagaa gaatggaaat attttatact gacagaaatc agtaatattt atatatttat	660
atttttaaaa tatttattta ttatatttata taagttcata ttccatattt attcaagatg	720
ttttaccgta ataattatta taaaaatat gcttct	756

<210> 9

<211> 825

<212> DNA

<213> Homo sapiens

<400> 9
atcactctct ttaatcacta ctcacattaa cctcaactcc tgccacaatg tacaggatgc 60
aactcctgtc ttgcattgca ctaattctt cacttgtcac aaacagtgc cctacttcaa 120
gttcgacaaa gaaaacaaaag aaaacacagc tacaactgga gcatttactg ctggatttac 180
agatgatttt gaatggaatt aataattaca agaatccaa actcaccagg atgctcacat 240
ttaagttta catgcccaag aaggccacag aactgaaaca gcttcagtgt ctagaagaag 300
aactcaaacc tctggaggaa gtgctgaatt tagctcaaag caaaaacttt cacttaagac 360
ccaggactt aatcagcaat atcaacgtaa tagttctgga actaaaggga tctgaaacaa 420
cattcatgtg tgaatatgc aatgagacag caaccattgt agaatttctg aacagatgga 480
ttaccttttgc tcaaagcattc atctcaacac taacttgata attaagtgt tcccacttaa 540
aacatatcag gccttctatt tatattttaa aatattttaa ttatattttt attgttgaat 600
gtatggttgc taccttattgt aactattatt cttaatctta aaactataaa tatggatctt 660
ttatgattct ttgttaagc cctagggct ctaaaatggc ttaccttatt tatccaaaaa 720
atatttatta ttatggttgaat tgttaaatat agtatctatg tagattggtt agtaaaaacta 780
tttaataat ttgataaata taaaaaaaaaaa aaacaaaaaaaaaa aaaaaa 825

<210> 10

<211> 15

<212> RNA

<213> Homo sapiens

<220>

<221> misc_feature

<222> (1)..(1)

<223> N = A, U, G, OR C

<220>

<221> misc_feature

<222> (15)..(15)

<223> N = A, U, G, OR C

<400> 10	
nauuuauuu uuuuu	15
<210> 11	
<211> 1125	
<212> DNA	
<213> Homo sapiens	
<400> 11	
ttctgcctc gagcccaccc ggaacgaaag agaagctcta tctcgccctcc aggagccccag	60
ctatgaactc cttctccaca agcgccctcg gtccagttgc cttctccctg gggctgctcc	120
tggtgttgcc tgctgccttc cctgccccag taccccccagg agaagattcc aaagatgttag	180
ccgccccaca cagacagcca ctcacccctt cagaacgaat tgacaaacaa attcggtaca	240
tcctcgacgg catctcagcc ctgagaaaagg agacatgtaa caagagtaac atgtgtgaaa	300
gcagcaaaga ggcactggca gaaaacaacc tgaaccttcc aaagatggct gaaaaagatg	360
gtatgttcca atctggattc aatgaggaga ctgccttggt gaaaatcatc actggcttt	420
tggagtttga ggtataccta gagtacctcc agaacagatt tgagagtagt gaggaacaag	480
ccagagctgt gcagatgagt acaaaagtcc tgatccagtt cctgcagaaa aaggcaaaga	540
atctagatgc aataaccacc cctgacccaa ccacaaatgc cagcctgctg acgaagctgc	600
aggcacagaa ccagtggctg caggacatga caactcatct cattctgcgc agctttaagg	660
agttccctgca gtccagcctg agggctttc ggcaaatgta gcatggcac ctcagattgt	720
tgttgttaat gggcattccct tcttctggtc agaaacctgt ccactggca cagaacttat	780
tttgtctct atggagaact aaaagtatga gcgtaggac actatttaa ttatttttaa	840
tttattaata tttaaatatg tgaagctgag ttaatttatg taagtcatat ttatatttt	900
aagaagtacc acttggaaaca ttttatgtat tagttttgaa ataataatgg aaagtggcta	960
tgcagtttga atatcctttt tttcagagcc agatcatttc ttggaaagtg taggcttacc	1020
tcaaataaat ggctaactta tacatattt taaagaata tttatattgt atttatataa	1080
tgtataaatg gtttttatac caataaatgg cattttaaaa aattc	1125
<210> 12	
<211> 3166	
<212> DNA	
<213> Homo sapiens	

<400> 12
 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagac gcgcgggcgt 60
 gcgagcagcg aaagcgacag gggcaaagtg agtgacctgc ttttgggggt gaccgccga 120
 gcgccgcgtg agccctcccc cttggatcc cgca gctgac cagtcgcgt gacggacaga 180
 cagacagaca ccccccag cccca gttac cacccctcc cccggccggcg gcggacagtg 240
 gacgcggcgg cgagccgcgg gcagggccg gagccgcgg ccggaggcgg ggtggagggg 300
 gtcggagctc gcggcgtcgc actgaaactt ttcttccaac ttctggctg ttctcgcttc 360
 ggaggagccg tgg tccgcgc ggggaaagcc gagccgagcg gagccgcgag aagtgc tagc 420
 tcggccggg aggagccgca gcggaggag ggggaggagg aagaagagaa ggaagaggag 480
 agggggccgc agtggcgact cggcgtcgg aagccgggt catggacggg tgaggcggcg 540
 gtgtgcgcag acagtgc tcc agcgcgcg cttccagcc ctggccggc ctggggccgg 600
 gaggaagagt agtcgcccga ggccgcgagg agagccggc gc cccac a cgc cgg 660
 gagggacgcg agccgcgcgc cccggcggg cctccgaaac catgaactt ctgtgtt 720
 ggtgcattt gggcatttgc ttgtgtctt acctccacca tgccaaatgg tccaggctg 780
 caccatggc agaaggagga gggcagaatc atcacgaatg ggtgaatgtc atggatgtct 840
 atcagcgcag ctactgccat ccaatcgaga ccctggtgga catcttccag gagtaccctg 900
 atgagatcga gtacatctt aagccatcct gtgtgc cctt gatgcgatgc ggggctgct 960
 ccaatgacga gggcctggag tgtgtccca ctgaggagtc caacatcacc atgcagatta 1020
 tgccggatcaa acctcaccaaa ggccagcaca taggagagat gagcttccata cagcacaaca 1080
 aatgtaatg cagaccaaag aaagatagag caagacaaga aaatccctgt gggccttgc 1140
 cagagccggag aaagcattt tttgtacaag atccgcagac gtgtaaatgt tcctgcaaaa 1200
 acacacactc gcgttgc aag gcgaggc a cttt gatc a cgt a cttc tgc a gat gtc 1260
 acaagccgag ggggtgagcc gggcaggagg aaggagccctc cctcagggtt tcgggaaacca 1320
 gatctctctc caggaaagac tgatacagaa c gat c gatac agaaaccacg ctggcccac 1380
 cacaccatca ccatcgacag aacagtcc tttccagaaa cctgaaatga aggaagagga 1440
 gactctgcgc agagcactt gggtccggag ggcgagactc cggcggaaac attccgggc 1500
 gggtgcacca gcacggtccc ttttggatt ggattcgcca ttttattttt cttgtgtcta 1560
 aatcaccgag cccggaaagat tagagat ttttctggg attccctgttag acacacccac 1620
 ccacatacat acat ttttat atatatat tatatatata taaaaataaa tatctctatt 1680
 ttatatatat aaaatata tatttttt ttaaatttaac agtgcataatg ttattgggt 1740
 cttcactgga tgtat tttgac tgctgtggac ttgagttggg agggaaatgt tccca ctcag 1800

atccctgacag ggaagaggag gagatgagag actctggcat gatcttttt ttgtcccaact
tggggggcc agggcctct cccctgccca agaatgtgca aggccagggc atgggggcaa
atatgaccca gtttggaa caccgacaaa cccagccctg gcgctgagcc tctctacccc
aggtcagacg gacagaaaga caaatcacag gttccggat gaggacacccg gctctgacca
ggagtttggg gagcttcagg acattgcgt gctttggga ttccctccac atgctgcacg
cgcatctcgc ccccagggc actgcctgga agattcagga gcctgggcgg ctttcgctta
ctctcacctg ctctgagtt gcccaggagg ccactggcag atgtcccggc gaagagaaga
gacacattgt tggaaagaagc agcccatgac agcgccccctt cctgggactc gccctcatcc
tcttcctgct ccccttcctg gggtgcagcc taaaaggacc tatgtcctca caccattgaa
accactagtt ctgtcccccc agggaaacctg gttgtgtgt tgtagtggt tgaccttcct
ccatccctg gtccctccct tccttcctg aggcacagag agacagggca ggatccacgt
gcccattgtg gaggcagaga aaagagaaaag tgtttatatacgttacgatcc ttaatatatcc
cttttaatt agaaattaga acagttaatt taatcaaaga gtagggtttt ttttcgtat
tcttggtaa tatttaattt caactattta tgagatgtat ctttgctct ctctgctct
cttatttgta ccgtttttg tatataaaat tcatgttcc aatctctc tcctgatcg
gtgacagtca cttagttatc ttgaacagat atttaattt gctaacaactc agctctgccc
tccccgatcc cctggctccc cagcacacat tccttgaaa gagggttca atatacatct
acatactata tatatatattgg gcaacttgta tttgtgtgta tatatatata tataatgtta
tgttatataatg tgatcctgaa aaaataaaaca tcgctattct gtttttata tgttcaaacc
aaacaagaaa aaatagagaaa ttctacatac taaatctc tccttttta attttaatat
ttgttatcat ttatatttg gtgctactgt ttatccgtaa taattgtggg gaaaagatata
taacatcacg tctttgtctc tagtgcagtt tttcgagata ttccgtagta catatttatt
tttaaacaac gacaaagaaa tacagatata tcttaaaaaa aaaaaaa 3166

<210> 13

<211> 249

<212> RNA

<213> Homo sapiens

<400> 13

ccggggcuau ggacggguga ggcggcgug ugcgccagaca gugcuccaggc gcgcgcgcuc

ccccagccccug gccccggccuc gggccgggag gaagaguagc ucggccgaggc gcccggaggaga 120

gcggggccgccc ccacagccccg agccggagag ggacgcgagc cgcgcgcccc ggucggggccu

ccgaaaccau gaacuuucug cugucuuggg ugcauuggag ccuugccuug cugcucuacc 240
uccaccaug 249

<210> 14

<211> 9181

<212> DNA

<213> Homo sapiens

<400> 14
ggtctctctg gtttagaccag atctgaggct gggagctctc tggctaacta gggAACCCAC 60
tgcttaagcc tcaataaaagc ttgccttgag tgcttcaagt agtgtgtgcc cgtctgttgt 120
gtgactctgg taactagaga tccctcagac ctttttagtc agtgtggaaa atctctagca 180
gtggcgccccg aacagggacc tgaaagcgaa agggaaacca gaggagctct ctcgacgcag 240
gactcggctt gctgaagcgc gcacggcaag aggcgagggg cgccgactgg tgagtacgcc 300
aaaaattttg actagcggag gctagaagga gagagatggg tgcgagagcg tcagtattaa 360
gcggggaga attagatcga tgggaaaaaaaaa ttccgttaag gccagggggaa aagaaaaaaaaat 420
ataaaattaaa acatatacgta tgggcaagca gggagctaga acgattcgcgtttaatcctg 480
gcctgttaga aacatcgaa ggctgttagac aaataactggg acagctacaa ccattcccttc 540
agacaggatc agaagaactt agatcattat ataatacagt agcaaccctc tattgtgtgc 600
atcaaaggat agagataaaaaa gacaccaagg aagcttttaga caagatagag gaagagcaaa 660
acaaaagtaa gaaaaaaagca cagcaagcag cagctgacac aggacacacgc aatcaggtca 720
gccaaaattta ccctatacgta cagaacatcc aggggcaaat ggtacatcg gccatatcac 780
ctagaacttt aaatgcatttgg gtaaaagtag tagaagagaa ggcttcagc ccagaagtga 840
tacccatgtt ttccgttta tcagaaggag ccacccaca agatttaaac accatgctaa 900
acacagtggg gggacatcaa gcagccatgc aaatgttaaa agagaccatc aatgagggaaag 960
ctgcagaatg ggatagatgtg catccagtgc atgcaggccc tattgcacca ggccagatga 1020
gagaaccaag gggaaatgtac atagcaggaa ctactagtac ctttcaggaa caaataggat 1080
ggatgacaaa taatccacccat atcccactgtg gagaattta taaaagatgg ataatcctgg 1140
gattaaataa aatagtaaga atgtatagcc ctaccagcat tctggacata agacaaggac 1200
caaagggacc cttagagac tatgttagacc ggttctataa aactctaaga gccgagcaag 1260
cttcacagga ggtaaaaaaaaat tggatgacag aaaccttggtt ggtccaaaat gcgAACCCAG 1320
attgttaagac tattttaaaaa gcattggac cagcggctac actagaagaa atgtatgacag 1380
catgtcaggc agtaggagga cccggccata aggcaagat tttggctgaa gcaatgagcc 1440

aagtaacaaa ttcagctacc ataatgatgc agagaggcaa ttttaggaac caaagaaaaga 1500
ttgttaagtg ttcaattgt ggcaaagaag ggcacacagc cagaaattgc agggccctta 1560
ggaaaaaggc ctgttgaaa tgtggaaagg aaggacacca aatgaaagat tgtactgaga 1620
gacaggctaa ttttttaggg aagatctggc cttcctacaa gggaggcca gggaaatttc 1680
ttcagagcag accagagcca acagccccac cagaagagag cttcagggtct gggtagaga 1740
caacaactcc ccctcagaag caggagccg tagacaagga actgtatcct ttaacttccc 1800
tcaggtcaact ctttggcaac gaccctcgat cacaataaag atagggggc aactaaagga 1860
agctctatta gatacaggag cagatgatac agtattagaa gaaatgagtt tgccaggaag 1920
atggaaacca aaaatgatag gggaaattgg aggtttatc aaagtaagac agtatgatca 1980
gatactcata gaaatctgtg gacataaagc tataggtaca gtatttagtag gacctacacc 2040
tgtcaacata attggaagaa atctgttgc tcagattggt tgcaactttaa atttcccat 2100
tagccctatt gagactgtac cagtaaaatt aaagccagga atggatggcc caaaagttaa 2160
acaatggcca ttgacagaag aaaaaataaa agcattagta gaaatttgc cagagatgg 2220
aaaggaaggg aaaatttcaa aaattggcc tgaaaatcca tacaatactc cagtattgc 2280
cataaagaaa aaagacagta ctaaatggag aaaattagta gatttcagag aacttaataa 2340
gagaactcaa gacttctggg aagttcaatt aggaatacc catccgcag gttaaaaaaa 2400
aaaaaaatca gtaacagtac tggatgtggg tcatgcatat tttcagttc ctttagatga 2460
agacttcagg aagtatactg catttaccat acctagtata aacaatgaga caccaggat 2520
tagatatcag tacaatgtgc ttccacaggg atggaaagga tcaccagcaa tattccaaag 2580
tagcatgaca aaaatcttag agccctttag aaaacaaaat ccagacatag ttatctatca 2640
atacatggat gatttgcatt taggatctga ctttagaaata gggcagcata gaacaaaaat 2700
agaggagctg agacaacatc tggtgaggtg gggacttacc acaccagaca aaaaacatca 2760
gaaagaacct ccattcttt ggtatgggta tgaactccat cctgataaat ggacagtaca 2820
gcctatagtg ctgccagaaa aagacagctg gactgtcaat gacatacaga agtttgtgg 2880
gaaattgaat tggcaagtc agatttaccc agggattaaa gtaaggcaat tatgtaaact 2940
ccttagagga accaaagcac taacagaagt aataccacta acagaagaag cagagctaga 3000
actggcagaa aacagagaga ttctaaaaga accagtacat ggagtgtatt atgaccatc 3060
aaaagactta atagcagaaa tacagaagca gggcaaggc caatggacat atcaaattta 3120
tcaagagcca tttaaaaatc tgaaaacagg aaaatatgca agaatgaggg gtgcccacac 3180
taatgtatgta aaacaattaa cagaggcagt gcaaaaaata accacagaaa gcatagtaat 3240
atggggaaag actcctaaat tttaactgccc catacaaaag gaaacatggg aaacatggtg 3300
gacagagtat tggcaagcca cctggattcc tgagtggag tttgttaata cccctccctt 3360

agtcaaatta tggtaaccagg tagagaaaga acccatagta ggagcagaaa ctttatgt	3420
agatgggca gctaacaggg agactaaatt aggaaaagca ggatatgtt ctaatagagg	3480
aagacaaaaa gttgtcaccc taactgacac aacaaatcg aagactgagt tacaagcaat	3540
ttatcttagt ttgcaggatt cgggattaga agtaaacata gtaacagact cacaatatgc	3600
attagaatc attcaagcac aaccagatca aagtgaatca gagttgtca atcaaataat	3660
agagcgtta ataaaaaagg aaaaggctta tctggcatgg gtaccagcac acaaaggaat	3720
tggagggaaat gaacaagtag ataaattagt cagtgttgc atcaggaaag tactatttt	3780
agatggaata gataaggccc aagatgaaca tgagaaatat cacagtaatt ggagagcaat	3840
ggctagtgtt tttaacctgc cacctgttgt agcaaaaagaa atagtagcca gctgtgataa	3900
atgtcagcta aaaggagaag ccatgcatgg acaagtagac tgtgtccag gaatatggca	3960
actagattgt acacatttag aaggaaaagt tatcctggta gcagttcatg tagccagtgg	4020
atataatagaa gcagaagtta ttccagcaga aacaggcag gaaacagcat attttctttt	4080
aaaatttagca ggaagatggc cagtaaaaac aatacatact gacaatggca gcaatttcac	4140
cggtgctacg gtagggccg cctgttggtg ggcgggaaatc aagcaggaat ttggattcc	4200
ctacaatccc caaatcagtaag gagtagttaga atctatgaat aaagaattaa agaaaattat	4260
aggacaggtta agagatcagg ctgaacatct taagacagca gtacaatgg cagtattcat	4320
ccacaatttt aaaagaaaag gggggattgg ggggtacagt gcagggaaa gaatagttaga	4380
cataatagca acagacatac aaactaaaga attacaaaaa caaattacaa aaattcaaaa	4440
tttccgggtt tattacaggg acagcagaaa tccactttgg aaaggaccag caaagctct	4500
ctggaaaggt gaaggggcag tagtaataca agataatagt gacataaaag tagtgccaaag	4560
aagaaaagca aagatcatta gggattatgg aaaacagatg gcaggtgtatg attgtgtggc	4620
aagttagacag gatgaggatt agaacatggaa aaagtttagt aaaacaccat atgtatgttt	4680
cagggaaagc tagggatgg tttatagac atcactatga aagccctcat ccaagaataa	4740
gttcagaagt acacatccc ctagggatg ctagattggt aataacaaca tattgggtc	4800
tgcatacagg agaaagagac tggcatttgg gtcagggagt ctccatagaa tggagggaaa	4860
agagatatacg cacacaagta gaccctgaac tagcagacca actaattcat ctgttattact	4920
ttgactgttt ttcagactct gctataagaa aggcottatt aggacacata gttagcccta	4980
ggtgtgaata tcaaggcagga cataacaagg taggatctct acaatacttg gcactagcag	5040
cattaataac accaaaaaaat ataaagccac ctttgcctag ttttacaaa ctgacagagg	5100
atagatggaa caagccccag aagaccaagg gccacagagg gagccacaca atgaatggac	5160
actagagctt ttagaggagc ttaagaatga agctgttaga cattttccta ggatttggct	5220
ccatggctta gggcaacata tctatgaaac ttatgggat acttggcag gagtggaaagc	5280

cataataaga attctgcaac aactgctgtt tatccatttt cagaattggg tgtcgacata 5340
gcagaatagg cgtaactcga cagaggagag caagaaatgg agccagtaga tcctagacta 5400
gagccctgga agcatccagg aagtcaagcct aaaactgctt gtaccaattt ctattgtaaa 5460
aagtgttgct ttcattgccca agtttggttc ataacaaaag ccttaggcat ctcctatggc 5520
aggaagaagc ggagacagcg acgaagagct catcagaaca gtcagactca tcaagcttct 5580
ctatcaaagc agtaagtagt acatgtaatg caacctatac caatagtagc aatagtagca 5640
ttagtagtag caataataat agcaatagtt gtgtggtcca tagtaatcat agaatatagg 5700
aaaatattaa gacaaagaaa aatagacagg ttaattgata gactaataga aagagcagaa 5760
gacagtggca atgagagtga aggagaaaata tcagcacttg tggagatggg ggtggagatg 5820
gggcaccatg ctccttggga tggatgtat ctgttagtgct acagaaaaat tggggtcac 5880
agtctattat ggggtacctg tggaaagga agcaaccacc actctatttt gtgcatacaga 5940
tgctaaagca tatgatacag aggtacataa tggggcc acacatgcct gtgtacccac 6000
agaccccaac ccacaagaag tagtatttgtt aaatgtgaca gaaaattttt acatgtggaa 6060
aaatgacatg gtggaaacaga tgcatacgata tataatcagt ttatggatc aaagcctaaa 6120
gccatgtgta aaattaaccc cactctgtgt tagttaaag tgcactgatt tgaagaatga 6180
tactaataacc aatagtagta gcgggagaat gataatggag aaaggagaga taaaaaactg 6240
ctcttcaat atcagcacaa gcataagagg taaggtgcag aaagaatatg cattttttta 6300
taaacttgat ataataccaa tagataatga tactaccagc tataagttga caagttgtaa 6360
cacctcagtc attacacagg cctgtccaaa ggtatccccc gagccattc ccatacatta 6420
ttgtgccccg gctggtttg cgattctaaa atgtataat aagacgttca atggAACAGG 6480
accatgtaca aatgtcagca cagtacaatg tacacatgga attaggccag tagtatcaac 6540
tcaactgtg ttaaatggca gtctagcaga agaagaggta gtaatttagat ctgtcaattt 6600
cacggacaat gctaaaacca taatagtaca gctgaacaca tctgttagaa ttaattgtac 6660
aagaccccaac aacaatacaa gaaaaagaat ccgtatccag agaggaccag ggagagcatt 6720
tgttacaata gaaaaatag gaaatatgag acaagcacat tgtaacatta gttagagcaa 6780
atggaaaac actttaaaac agatagctag caaattaaga gaacaattt gaaataataa 6840
aacaataatc tttaaagcaat cctcaggagg ggacccagaa attgtaacgc acagttttaa 6900
ttgtggaggg gaatttttct actgttaattc aacacaactg tttatagta cttggttttaa 6960
tagtacttgg agtactgaag ggtcaaataa cactgaagga agtgacacaa tcaccctccc 7020
atgcagaata aaacaaattha taaacatgtg gcagaaagta ggaaaagcaa tggatggccc 7080
tcccatcgtt ggacaaattha gatgttcatc aaatattaca gggctgtat taacaagaga 7140
tggtggtat agcaacaatg agtccgagat cttcagaccc ggaggaggat atatgaggga 7200

caattggaga agtgaattat ataaatataa agtagtaaaa attgaaccat taggagtagc	7260
acccaccaag gcaaagagaa gagtggtgca gagagaaaaa agagcagtgg gaataggagc	7320
tttgttcctt gggttcttgg gagcagcagg aagcactatg ggccgcgcct caatgacgct	7380
gacgg tacag gccagacaat tattgtctgg tatagtgcag cagcagaaca atttgctgag	7440
ggctattgag ggcacacagg atctgttgc actcacagtc tggggcatca agcagctcca	7500
ggcaagaatc ctggctgtgg aaagataacct aaaggatcaa cagtccttgg ggatttgggg	7560
ttgctcttggaa aaactcattt gcaccactgc tgtgccttgg aatgcttagtt ggagtaataa	7620
atctcttggaa cagatttggaa atcacacgc acggatggag tggacagag aaattaacaa	7680
ttacacaagc ttaatacact ccttaatttga agaatcgcaa aaccagcaag aaaagaatga	7740
acaagaatta ttgaaatttag ataaatggc aagtttggg aattggtttta acataacaaa	7800
ttggctgtgg tatataaaaat tattcataat gatagtagga ggcttggtag gtttaagaat	7860
agttttgct gtactttcta tagtgaatag agtaggcag ggatattcac cattatcggt	7920
tcagacccac ctccccacccc cgaggggacc cgacaggccc gaaggaatag aagaagaagg	7980
tggagagaga gacagagaca gatccattcg attagtgaac ggatccttgg cacttatctg	8040
ggacgatctg cggagcctgt gccttccatcg ctaccaccgc ttgagagact tactcttgat	8100
tgtaacgagg attgttggaaac ttctggacg caggggggtgg gaagccctca aatattggtg	8160
gaatctccta cagtatttggaa gtcaggaact aaagaatagt gctgttagt tgctcaatgc	8220
cacagccata gcagtagctg agggacaga tagggttata gaagtagtac aaggagcttgc	8280
tagagctatt cgccacatac cttagaagaat aagacagggc ttggaaagga ttttgctata	8340
agatgggtgg caagtggtaaaaatg tgattggatg gcctactgtt aaggaaagaa	8400
tgagacgagc tgagccagca gcagataggg tggagcagc atctcgagac ctggaaaaac	8460
atggagcaat cacaagtagc aatacagcag ctaccaatgc tgcttgcgcc tggctagaag	8520
cacaagagga ggaggaggtt ggtttccag tcacacctca ggtacccat agaccaatga	8580
cttacaaggc agctgttagat cttagccact tttaaaaga aaaggggggc ctggaaggc	8640
taattcactc ccaaagaaga caagatatcc ttgatctgtg gatctaccac acacaaggct	8700
acttccctga ttagcagaac tacacaccag ggcgcagggtt cagatatcca ctgacccat	8760
gatggtgcta caagcttagt ccagttgagc cagataagat agaagaggcc aataaaggag	8820
agaacaccag ctgttacac cctgtgagcc tgcattggat ggatgaccccg gagagagaag	8880
tgttagagtg gaggtttgc acggcccttag catttcataca cgtggcccgagcgtgcac	8940
cggagtactt caagaactgc tgacatcgag ctgttacaa gggacttcc gctggggact	9000
ttccaggag gctgtggcctg ggccggactg gggagtgccg agccctcaga tcctgcata	9060
aagcagctgc tttttgcctg tactgggtct ctctggtag accagatctg agcctggag	9120

ctctctggct aactaggaa cccactgctt aagcctcaat aaagcttgcc ttgagtgctt 9180

c 9181

<210> 15

<211> 29

<212> RNA

<213> Homo sapiens

<400> 15

ggcagaucug agccugggag cucucugcc 29

<210> 16

<211> 52

<212> RNA

<213> Homo sapiens

<400> 16

uuuuuuuagg aagaucuggc cuuccuacaa gggaggcca gggauuuuuc uu 52

<210> 17

<211> 9413

<212> DNA

<213> Homo sapiens

<400> 17

ttggggcgca cactccacca tagatcactc ccctgtgagg aactactgtc ttcacgcaga 60

aagcgtctag ccatggcggtt agtatgagtg ttgtgcagcc tccaggaccc cccctcccg 120

gagagccata gtggctcgcg gaaccggta gtacaccgga attgccagga cgaccgggtc 180

ctttcttggta tcaacccgtt caatgcctgg agatttgggc gtgccccgc gagactgtta 240

gcccggtagt gttgggtcgc gaaaggcctt gtggtaactgc ctgataggggt gcttgcgagt 300

gccccggag gtctcgtaga ccgtgcatca tgagcacaaa tcctaaacct caaaagaaaa 360

ccaaacgtaa caccaaccgc cgccccacagg acgttaagtt cccggggcggt ggtcagatcg 420

ttgggtggagt ttacctgttg ccgcgcaggc gccccaggtt gggtgtgcgc gcgacttagga 480

agacttccga gcggtcgcaa cctcgtggaa ggcgacaacc tatccccaaag gctcgccggc 540

ccgagggttag gacctgggtt cagccgggtt acccttggcc cctctatggc aacgaggta 600

tggggtgtggc	aggatggctc	ctgtcacccc	gtggctctcg	gcctagttgg	ggcccccacag	660
accccccggcg	taggtcgctgt	aatttggta	aggtcatcga	tacccttaca	tgcggcttcg	720
ccgacctcat	gggtacatt	ccgcttgtcg	gcgcxxxxct	agggggcgct	gccagggccc	780
tggcacatgg	tgtccgggtt	ctggaggacg	gcgtgaacta	tgcaacaggg	aatctgccc	840
gttgctctt	ctctatcttc	ctcttagctt	tgctgtcttg	tttaccatc	ccagttccg	900
cttacgaggt	gwgcaacgtg	tccggatat	accatgtcac	gaacgactgc	tccaactcaa	960
gtatttgtta	tgaggcagcg	gacatgatca	tgcacacccc	cgggtgcgtg	ccctgcgtcc	1020
gggagagtaa	tttctccgt	tgctggtag	cgctcactcc	cacgctcg	gccaggaaca	1080
gcagcatccc	caccacgaca	atacgacgccc	acgtcgattt	gctcggtgg	gcggctgctc	1140
tctgttccgc	tatgtacgtt	ggggatctct	gccccatccgt	ttttctcg	tccagctgt	1200
tcaccttctc	acctcgccgg	tatgagacgg	tacaagattt	caattgctca	atctatcccg	1260
gccacgtatc	aggtcacccgc	atggcttggg	atatgatgat	gaactggtca	cctacaacgg	1320
ccctagtggt	atcgagcta	ctccggatcc	cacaagccgt	cgtggacatg	gtggcggggg	1380
cccactgggg	tgtcctagcg	ggccttgcc	actattccat	ggggggaaac	tgggctaagg	1440
tcttgattgt	gatgctactc	tttgcggcg	ttgacggca	caccacgtg	acagggggaa	1500
gggtgcctc	cagcacccag	agcctcg	cctggctctc	acaaggccca	tctcagaaaa	1560
tccaaactcg	gaacaccaac	ggcagctggc	acatcaacag	gaccgctctg	aattgcaatg	1620
actccctcca	aactgggttc	attgctgc	tgttctacgc	acacagg	aacgcgtccg	1680
ggtgcccaga	gocatggct	agctgccc	ccatcgatga	gttcgctc	gggtgggg	1740
ccatcaactca	tgatatgcct	gagagctcg	accagaggcc	atattgctgg	caaacgcgc	1800
ctcgaccgtg	cgggatcg	cctgcgtc	agggtgtgg	tccagtgtat	tgcttca	1860
cgagccctgt	tgtatgtggg	acgaccgatc	gttccggc	tcctacgtat	agctgggggg	1920
agaatgagac	agacgtgt	ctacttagca	acacgcggcc	gcctcaaggc	aactggttt	1980
ggtgcacgtg	gatgaacagc	actgggtca	ccaagacgtg	cggggccct	ccgtgcaaca	2040
tcgggggggt	cggcaacaac	accttggtct	gccccacgg	ttgcttccgg	aagcaccccg	2100
aggccactta	cacaaagtgt	ggctcgggc	cctgggtgac	acccagg	tggttgact	2160
accatacag	gctctggcac	tacccctgca	ctgttaactt	tacccgtt	aaggtcagga	2220
tgtatgtggg	gggcgtggag	cacaggctca	atgctgc	caattggact	cgaggagagc	2280
gctgtgactt	ggaggacagg	gatagg	actcagccc	gctgtgtcg	tctacaacag	2340
agtggcagat	actgcctgt	tccttacca	ccctaccgg	cctgtccact	ggcttgc	2400
atcttcaccg	gaacatcg	gacgtgcaat	acctgtacgg	tatagg	gcagttgtct	2460
ccttgcaat	caaatggag	tatatcctgt	tgctttcc	tcttgcgg	gacgcgcg	2520

tctgtgcctg	cttgtggatg	atgctgtga	tagcccaggc	tgaggccacc	ttagagaacc	2580
tggtgtcct	caatgcggcg	tctgtggccg	gagcgcatgg	ccttctctcc	ttcctcgtgt	2640
tcttcgtcgc	cgcctggta	atcaaaggca	ggctggtccc	tggggcgca	tatgctctct	2700
atggcgtatg	gcccgttgc	ctgctcttgc	tggcattacc	accacgagct	tatgccatgg	2760
accgagagat	ggctgcatcg	tgcggaggcg	cggtttttgt	aggctggta	ctcttgacct	2820
tgtcaccata	ctataagggt	ttcctcgcta	ggctcatatg	gtggttacaa	tattttatca	2880
ccagagccga	ggcgcacttg	caagtgtggg	tccccctct	aatgttcgg	ggaggcccg	2940
atgccatcat	cctccttaca	tgcgcggtcc	atccagagct	aatctttgac	atcaccaaac	3000
tcctgctcgc	catactcggt	ccgctcatgg	tgcctccaggc	tggcataact	agagtgccgt	3060
actttgtacg	cgcgcagggg	ctcatccgt	catgcatgtt	agtgcggaag	gtcgctggag	3120
gccactatgt	ccaaatggcc	ttcatgaagc	tggccgcgt	gacaggtacg	tacgtatatg	3180
accatcttac	tccactgcgg	gattgggccc	acgcgggcct	acgagacett	gcgggtggcag	3240
tagagccgt	cgtcttctct	gacatggaga	ctaaactcat	cacctggggg	gcagacaccg	3300
cggcgtgtgg	ggacatcatc	tcgggtctac	cagtctccgc	ccgaaggggg	aaggagatac	3360
ttcttaggacc	ggccgatagt	tttggagagc	aggggtggcg	gctccttgcg	cctatcacgg	3420
cctattccca	acaaacgcgg	ggcctgttt	gctgtatcat	cactagcctc	acaggtcggg	3480
acaagaacca	ggtcgatggg	gaggttcagg	tgcctccac	cgcaacgcaa	tcttcctgg	3540
cgacctgcgt	caatggcgt	tgttggaccg	tctaccatgg	tgcggctcg	aagaccctgg	3600
ccggcccgaa	gggtccaatc	acccaaatgt	acaccaatgt	agaccaggac	ctcgctggct	3660
ggccggcgcc	ccccggggcg	cgctccatga	caccgtgcac	ctgcggcagc	tcggaccttt	3720
acttggtcac	gaggcatgt	gatgtcggtc	cggtgcccg	gcggggcgac	agcaggggg	3780
gcctgctttc	ccccaggccc	atctcttacc	tgaagggtctc	ctcggtgga	ccactgcttt	3840
gcccttcggg	gcacgttgta	ggcatcttcc	ggcgtgtgt	gtgcacccgg	ggggttgcga	3900
aggcggtgga	cttcataaccc	gttgagtcta	tggaaactac	catgcgtct	ccggcttca	3960
cagacaactc	atcccctccg	gccgtaccgc	aaacattcca	agtggcacat	ttacacgctc	4020
ccactggcag	cgcaagagc	accaaagtgc	cggctgcata	tgcagccaa	gggtacaagg	4080
tgctcgtcct	aaacccgtcc	gttgccgcca	cattgggctt	tggagcgtat	atgtccaagg	4140
cacatggcat	cgagcctaac	atcagaactg	ggtaaggac	catcaccacg	ggcggcccc	4200
tcacgtactc	cacctattgc	aagttccttgc	ccgacggtg	atgctccggg	gggcctatg	4260
acatcataat	atgtgatgaa	tgccactcaa	ctgactcgac	taccatcttgc	ggcgtcgca	4320
caagtcccttgc	tcaggcagag	acggctggag	cgcggctcg	cgtgctcgcc	accgcccacgc	4380
ctccgggatc	gatcaccgt	ccacacccca	acatcgagga	agtggccctg	tccaaacactg	4440

gagagattcc	cttcttatggc	aaagccatcc	ccattgaggc	catcaagggg	ggaaggcatc	4500
tcatttctg	ccattccaag	aagaagtgt	acgagctcg	cgc当地	acaggcctcg	4560
gactcaatgc	tgttagcgtat	taccgggtc	tcgatgtgtc	cgtcataccg	actagcggag	4620
acgtcggtgt	cgtggcaaca	gacgctctaa	tgacgggtt	taccggcgac	tttgactcag	4680
tgatcgactg	caacacatgt	gtcaccaga	cagtcgattt	cagcttggat	ccccacccatca	4740
ccattgagac	gacaacgctg	ccccaaagacg	cgggtcg	tgc当地	cgaggttagga	4800
ctggcagggg	caggagtggc	atctacaggt	tttgactcc	aggagaacgg	ccctcaggca	4860
tgttcactc	ctcggtcctg	tgtgagtgt	atgacgcagg	ctgc当地	tatgagctca	4920
cgc当地	gacctcggtt	aggttgcggg	cttacctaaa	tacaccagg	ttgcccgtct	4980
gccaggacca	ccttagatcc	tggagagcg	tcttacagg	cctcacccac	atagatgccc	5040
acttcttgc	ccagacccaaa	caggcaggag	acaacctccc	ctacctggta	gcataccaag	5100
ccacagtgt	cggcagggt	caggctccac	ctccatcg	ggaccaaatg	tggaagtgtc	5160
tcatacggt	aaagcccaca	ctgcatggc	caacgcccct	gctgtacagg	ctaggagccg	5220
ttcaaaatga	ggtcactctc	acacacccca	taaccaaata	catcatgg	tgcatgtcgg	5280
ctgacctgga	ggtcgtact	agcacctgg	tgcttagtag	cggagtcctt	gc当地	5340
ccgc当地	cctgacgaca	ggcagcgtgg	tcattgtgg	caggatcatc	ttgtccggg	5400
ggccagctgt	tattcccgac	agggaaatcc	tctaccagga	gttcgatgag	atgaaagagt	5460
gtgcttcaca	cotcccttac	atcgagcaag	aatgcagct	cggc当地	ttcaaacaga	5520
aggcgctcg	attgctgca	acagccacca	agcaagcgg	ggctgctgt	cccgtgg	5580
agtccaaatg	gcaagccctt	gaggtttct	gggc当地	catgtgg	ttcatcagcg	5640
ggatacagta	cttggcaggc	ctatccactc	tgctggaaa	ccccgc当地	gcatcattga	5700
tggctttac	agcctctatc	accagccgc	tcaccaccc	aaataccctc	ctgtttaaaca	5760
tcttgggggg	atgggtggct	gcccaactcg	ctccccccag	cgctgcttcg	gcttcgtgg	5820
gc当地	tgccgggtcg	gccgtggca	gcatagg	cggaaaggta	tttgtggaca	5880
ttctggcggg	ctatggggcg	gggggtggct	gcmcactcg	ggc当地	gtcatgagcg	5940
gc当地	ctccactgag	gatctggta	atttactccc	tgccatcctt	tctctggcg	6000
ccctgggtgt	cgggggtcg	tgcgaccaa	tactgcgtcg	gc当地	ccgggagagg	6060
gggctgtgca	gtggatgaac	cggctgatag	cgttcgcttc	gc当地	cacgtctccc	6120
ccacgcacta	tgtccccgag	agcgacgccc	cggcgcgtgt	tactcagatc	ctctccagcc	6180
ttaccatcac	tca	tttgcgt	aaaggatgtt	gggactggat	atgc当地	6240
cttggccgg	ctcggtggcta	aaggatgtt	gggactggat	atgc当地	ttgagtgact	6300
tcaagacttg	gctccacttcc	aagctccatgc	cgc当地	gggactccct	ttcctgtcat	6360

gccaacgcgg	gtacaaggga	gtctggcggg	ggatggcat	catgcaaacc	acctgcccatt	6420
gtggagcaca	gatcacccgga	catgtcaaaa	atggctccat	gaggattgtt	ggccaaaaaa	6480
cctgcagcaa	cacgtggcat	ggaacattcc	ccatcaacgc	atacaccacg	ggcccttgca	6540
cgcctcccc	agcgccgaac	tattccaggg	cgctgtggcg	ggtggctgct	gaggagtacg	6600
tggaggttac	gccccgtgggg	gatttccact	acgtgacggg	catgaccact	gacaacgtga	6660
aatgcccatt	ccaggttcca	gcccctgaat	tttcacgga	ggtggatgga	gtacggttgc	6720
acaggtatgc	tccagtgtgc	aaacctctcc	tacgagagga	ggtcgtattc	caggtcgggc	6780
tcaaccagta	cctggtcggg	tcacagctcc	catgtgagcc	cgaaccggat	gtggcagtgc	6840
tcacttccat	gctcaccgac	ccctctcata	ttacagcaga	gacggccaag	cgtaggctgg	6900
ccaggggtc	tcccccctcc	ttggccagct	cttcagctag	ccagttgtct	gccccttctt	6960
tgaaggcgac	atgtactacc	catcatgact	ccccggacgc	tgacctcatc	gaggccaacc	7020
tcctgtggcg	gcaggagatg	ggcgggaaca	tcacccgtgt	ggagtcagaa	aataaggtgg	7080
taatcctgga	ctctttcgat	ccgattcggg	cggtggagga	tgagaggaa	atatccgtcc	7140
cggcggagat	cctgcgaaaaa	cccaggaagt	tcccccagc	gttgc当地	tggcacgcc	7200
cggattacaa	ccctccactg	ctagagtcc	ggaaggaccc	ggactacgtc	ccccgggtgg	7260
tacacgggtg	ccctttgcca	tctaccaagg	cccccccaat	accacctcca	cgaggaaaga	7320
ggacggttgt	cctgacagag	tccaccgtgt	cttctgcctt	ggcggagctc	gctactaaga	7380
cctttggcag	ctccgggtcg	tcggccgttg	acagcggcac	ggcactggc	cctcccgatc	7440
aggcctccga	cgacggcgac	aaaggatccg	acgttgagtc	gtactcetcc	atgccccccc	7500
tcgagggaga	gccaggggac	cccgaccta	gacggggtc	ttggtctacc	gtgagcgggg	7560
aagctggtga	ggacgtcgtc	tgctgctcaa	tgcctatac	atggacaggt	gccttgcata	7620
cgcctgcgc	tgccggaggag	agcaagttgc	ccatcaatcc	gttgc当地	tcttgctgc	7680
gtcaccacag	tatggtctac	tccacaacat	ctcgacgcgc	aagtctgcgg	cagaagaagg	7740
tcaccttga	cagactgcaa	gtcctggacg	accactaccg	ggacgtgctc	aaggagatga	7800
aggcgaaggc	gtccacagtt	aaggcttaggc	ttcttatctat	agaggaggcc	tgcaaactga	7860
cgcctccaca	ttcggccaaa	tccaaatttg	gctacggggc	gaaggacgtc	cggagcctat	7920
ccagcagggc	cgtcaaccac	atccgcctcg	tgtggagga	cttgc当地	gacactgaaa	7980
caccaattga	taccaccatc	atggcaaaaaa	atgaggttt	ctgcgtccaa	ccagagaaag	8040
gaggccgcaa	gccagctcg	tttatcgat	tccagaccc	gggggtacgt	gtatgcgaga	8100
agatggccct	ttacgacgtg	gtctccaccc	ttcctcaggc	cgtgatggc	ccctcatacg	8160
gattccagta	ctctcctggg	cagcgggtcg	agttcctggt	gaatacctgg	aaatcaaaga	8220
aatgccttat	ggcattctca	tatgacaccc	gctgcttga	ctcaacggtc	actgagaatg	8280

acatccgtac tgaggaatca atttaccaat gttgtgactt ggcccccaa gccaggcagg	8340
ccataaggc gctcacagag cggcttatg tcgggggtcc cctgactaat tcgaaggggc	8400
agaactgcgg ttatcgccgg tgccgcgeaa gtggcgtgct gacgactagc tgcggcaaca	8460
ccctcacatg ttacttgaag gccactgcgg cctgtcgagc tgcaaagctc caggactgca	8520
cgtatgcgtgtaacggagac gaccttgctg ttatctgtga gagtgcggga acccaggagg	8580
atgcggcggc cctacgagcc ttcacggagg ctatgactag gtattccgcc cccccgggg	8640
accgcggcca accagaatac gacttggagc tgataacgctc atgctcctcc aatgtgtcg	8700
tcgcgcacga tgcatccggc aaaagggtgt actacctcac ccgtgacccc accacccccc	8760
tcgcacgggc tgcgtggag acagtttagac acactccagt caactcctgg ctaggcaata	8820
tcatcatgta tgcccacc ctagggcga ggatgattct gatgactcat ttcttccta	8880
tccttc tagc tcaggagcaa cttgaaaaag ccctggatttgc ttagatctac gggcgtt	8940
actccattga gccacttgac ctacccaga tcattgaacg actccatggt cttagcgcac	9000
tttcactcca cagttactct ccaggtgaga tcaatagggt ggcttcatgc ctcaggaaac	9060
ttggggtacc gccttgcga gtctggagac atcggccag aagtgtccgc gctaagctac	9120
tgtcccaggg ggggagggtt gcaacttgcg gcaagtacactt cttcaactgg gcagtaaaga	9180
ccaagctaa actcactcca atcccgctg cgtcccagct agacttgtcc ggctggttcg	9240
ttgctggta caacggggga gacatatatc acagcctgca tcgtgccga ccccggtt	9300
tcatgttgc tctactccta ctttctgttag gggtaggcat ctacctgctc cccaaaccgg	9360
gaacggggag ctaaccactc caggccaaata ggccattccc ttttttttt ttc	9413

<210> 18

<211> 328

<212> RNA

<213> Homo sapiens

<400> 18	
uuggggcga cacuccacca uagaucacuc cccugugagg aacuacuguc uucacgcaga	60
aagcgucuag ccauggcguu aguaugagug uugugcagcc uccaggaccc ccccuccccgg	120
gagagccaua guggucugcg gaacctgguga guacaccgga auugccagga cgaccggguc	180
cuuucuugga ucaaccgcu caaugccugg agauuugggc gugcccccgc gagacugcua	240
gccgaguagu guugggucgc gaaaggccuu gugguacugc cugauagggu gcuugcggagu	300
gccccgggag gucucguaga ccgugcau	328

<210> 19
<211> 14
<212> RNA
<213> Homo sapiens

<400> 19
auuuugggcgu gccc

14

<210> 20
<211> 27
<212> RNA
<213> Homo sapiens

<400> 20
gccgaguagu guuggggucgc gaaaggc

27

<210> 21
<211> 340
<212> DNA
<213> Homo sapiens

<400> 21
atggcgag ggaagctcat cagtggggcc acgagctgag tgcgtcctgt cactccactc 60
ccatgtccct tgggaaggtc tgagactagg gccagaggcg gccctaacag ggctctccct 120
gagcttcagg gaggtgagtt cccagagaac ggggctccgc gcgaggtcag actgggcagg 180
agatgccgtg gaccccgccc ttccccggagg ggccccggcgg atgcctcctt tgccggagct 240
tggaacacac tcacggccag cgaagtgagt tcaatggctg aggtgaggta ccccgcagg 300
gacctcataa cccaaattcag accactctcc tccgcccatt 340

<210> 22
<211> 349
<212> DNA
<213> Homo sapiens

<400> 22

gaggaaagtc cgggctcaca cagtctgaga tgattgtagt gttcgtgctt gatgaaacaa	60
taaatcaagg cattaatttgc acggcaatga aatatcctaa gtcttcgat atggatagag	120
taatttggaaa gtgccacagt gacgttagtt ttatagaaat ataaaagggtg gaacgcggta	180
aaccctcgatgtgagcaatc caaatttggt aggagcactt gtttaacgga attcaacgta	240
taaacgagac acacttcgcaaatgaagtgtgttagacag atggttatca cctgagtacc	300
agtgtgacta gtgcacgtga tgagtacgt ggaacagaac gcggcttat	349

<210> 23

<211> 377

<212> DNA

<213> Homo sapiens

<400> 23	
gaagctgacc agacagtcgc cgcttcgtcg tcgtcctt cgggggagac gggcggaggg	60
gaggaaagtc cgggctccat agggcagggt gccaggtaac gcctgggggg gaaaccccacg	120
accagtgc当地 cagagagcaa accgcgcatg gccgcgc当地 gcgggatcag gtaagggtga	180
aagggtgc当地 taagagcgca ccgc当地ggct ggtaacagtc cgtggcacgg taaactccac	240
ccggagcaag gccaaatagg gttcataag gtacggcccg tactgaaccc gggtaggctg	300
cttgagccag tgagcgattt ctggctaga tgaatgactg tccacgacag aaccggctt	360
atcggtcagt ttcacct	377

<210> 24

<211> 38110

<212> DNA

<213> Homo sapiens

<400> 24	
ccaccggta cgatcttgc gaccatggcc ccacaatagg gccggggaga cccggcgtca	60
gtggggcg gcacggtc当地 taacgtctgc gcaacacggg gttgactgac gggcaatatc	120
ggctccatag cgtc当地ggccgc ggatacagta aaggagcatt ctgtgacggaa aagacgccc	180
gacgacgtct tcaaacttgc caaggacgag aaggctgaat atgtgacgt cccgttctgt	240
gacctgc当地tgc gcatcatgca gcacttcacg attccggctt cggc当地ttga caagagcgtg	300
tttgacgacg gcttggc当地t tgacggctcg tggattcgc当地 ggttccagtc gatccacgaa	360
tccgacatgt tgcttcttcc cgatcccgag acggcgc当地ca tcgaccgctt cccgc当地ggcc	420

aagacgctga atatcaactt ctttgtcac gaccgttca ccctggagcc gtactcccgc	480
gaccgcgca acatcgccc caaggccgag aactacctga tcagcactgg catcgccgac	540
accgcatact tcggcgccga ggccgagttc tacatttcg attcggttag cttcgactcg	600
cgcgccaacg gctccttcta cgaggtggac gccatctcg gggtggaa caccggcg	660
gcgaccgagg ccgacggcag tcccaaccgg ggctacaagg tccgccacaa gggcggtat	720
ttcccagtgg cccccaacga ccaatacgac gacctgcgcg acaagatgt gaccaacctg	780
atcaactccg gttcatcct ggagaaggc caccacgagg tggcagcgg cgacaggcc	840
gagatcaact accagttcaa ttgcgtctg cacgcccgg acgacatgca gttgtacaag	900
tacatcatca agaacaccgc ctggcagaac ggcaaaacgg tcacgttcat gccaagccg	960
ctgttcggcg acaacgggtc cgccatgcac tgtcatcagt cgctgtggaa ggacggggcc	1020
ccgctgatgt acgacgagac gggttatgcc ggtctgtcg acacggcccg tcattacatc	1080
ggcggctgt tacaccacgc gccgtcgctg ctggcattca ccaacccgac ggtgaactcc	1140
tacaagcggc tggttccggg ttacgaggcc ccgatcaacc tggctatag ccagcgcaac	1200
cggtcggcat gcgtgcgcatt cccgatcacc ggcaagcaacc cgaaggccaa gcggctggag	1260
ttccgaagcc ccgactcgtc gggcaaccgg tatctggcgt tctcgccat gctgatggca	1320
ggcctggacg gtatcaagaa caagatcgag ccgcaggcgc ccgtcgacaa ggatctctac	1380
gagctgccgc cgaaagaggc cgcgagtatc ccgcagactc cgacccagct gtcagatgt	1440
atcgaccgtc tcgaggccga ccacgaatac ctcaccgaag gaggggtgtt cacaacgac	1500
ctgatcgaga cgtggatcag ttcaagcgc gaaaacgaga tcgagccggt caacatccgg	1560
ccgcatccct acgaattcgc gctgtactac gacgtttaag gactctcgc agtccgggt	1620
tagagggagc ggcgtgtcgt tgccaggcg ggcgtcgagg ttttcgatg ggtgacgggt	1680
gccggcaacg ggcgcggcgc acccgctcg aagagccgt ttaagaacgt tcaaggacgt	1740
ttcagccggg tgccacaacc cgcttgcaa tcatctcccg accgcccggc gggttgtctt	1800
tcacatgcgc cgaaactcaa gccacgtcgt cgcccaggcg tgcgtcgac gcccgttcag	1860
gttaagtgtc ggggattcgt cgtcgccgc ggcgtccacg ctgaccaacg gggcagtcaa	1920
ctcccaaca ctttgcgcac taccgcctt gcccggcg tcaccgtag gtagttgtcc	1980
aggaattccc caccgtcgtc gtttgcgcag cccggccgcga ccgcgaccgc attgagctgg	2040
cggccgggtc cggcagctg gtcggggc ttgcggcgca ccaacaccag cgcggtcg	2100
gcccgggtgg cggtcagcca ggcgtacgg agcagctcca cgtcggtgc gggaccaga	2160
tccggccgcg cgatgacatc cagggattgc agcgtcgagg tgggtgcag ggcgggaacc	2220
tggtgccat gctgttagctg cagcaactgc acggtccatt cgatgtcgcc cagtccgccc	2280
cggcccaagtt tggtgtgtgt gttgggtcg gcaccgcgc gcaaccgcgc ggactcgata	2340

cgggccttga	tgcggcgaat	ctcgccacc	gagtcagcgg	acacaccgtc	gggcggatac	2400	
cgcgtttgt	cgaccatccg	taggaatcgc	tgacccaact	cggcatcgcc	ggcaaccgcg	2460	
tgtgcgcgt	ta	gcagggcctg	gatctccat	ggtgtgccc	actgctcgta	gtatgcggcg	2520
taggacccc	gggtgcggac	cagcggaccg	ttgcggccct	cgggtcgcaa	attggcgtcg	2580	
agctccagcg	gcccgtcgac	gctgggtgtc	cccaggcagcg	cccgaaacccg	ctcgccgatc	2640	
gtatgcgacc	atttcaccgc	ccgtgcac	tcgacgcgg	tggccggctc	acagacgaac	2700	
atcacgtcgg	catccgaccc	gtagccaaac	tcggcaccac	ccagccgacc	catgccatg	2760	
accgcgatgg	ccgccccggg	gcgatcg	tcgggaaggc	tggcccgat	catgacgtcc	2820	
agcgccgcct	gcagcaccgc	cacccacacc	gacgtcaacg	ccggcacac	ctcggtgacc	2880	
tcgagcaggc	cgagcagg	cgccgaaccg	atgcggcca	gctctcgacg	acgcagcgtg	2940	
cgcgcccg	cgatggcccg	ctccgggtcg	ggtagcg	tcgcccggc	gatcagcgcc	3000	
cgagccacgg	cgccgggctc	ggtctcgagc	agttcggc	ccgcaggccc	gtcctcgat	3060	
tgctggatga	cccgccgcgc	g	gcgcataac	agatccggca	catacgccga	3120	
acatgcatga	ccgcgttggc	caccgcggg	ttgtccc	gcgtggccag	gtaccagctt	3180	
tcggtgccca	gcgcctcact	gagccgcgg	taggcagc	gtccgcgtc	ggatcgggg	3240	
gcatacgaca	tccagtccag	cagcctggc	agcagcaccg	actgcaccc	tccgcgcgg	3300	
ccgcgttgc	tgaccaacgc	cgacatgtgt	ttcaacgcgg	tctgcgg	ctcgtagccc	3360	
agcgcggcca	gccggcgccc	cgcggctcc	aacgtcatgc	cgtggcgat	ctccaacccg	3420	
gtcggccga	tcgattccag	cagcgg	ttgaagagtt	tggtgtgtaa	cttcgacacc	3480	
cgcacgttct	gttcttgag	ttcctccgc	agcaccc	ccgcatacg	tcggccatcg	3540	
ggccggatgt	ggccgcgc	cgccagccag	cgca	cctcg	ggatcggg	3600	
agcaggtggg	tgcgcttgc	ccgctgcaac	tgca	tcgagcag	cctgaggaac	3660	
tcatacgacg	cggtcatgtt	cgccgcgtcc	tcacgccc	tgtagcc	ttcgccc	3720	
gccgccaatg	cgtccaccgt	ggacgc	cgtaacgact	cgtcg	ggcatgaacc	3780	
agctgcagta	gctgtacggc	gaactccacg	tcgc	cgccg	gagttgagc	3840	
tcgcggccgc	ggacatcg	ggcacc	tg	ccaccc	ccgcgc	3900	
tcgaccacaa	agtcttcgc	ctcg	cagg	ccacca	ggcggtcagg	3960	
taacgctcgc	caagttccgc	gtcg	ccaa	ctt	cagcaa	4020	
tcccagg	tgc	ccatgc	actggccgt	c	cgctgaaac	4080	
tcccagg	gc	ccatgc	actggccgt	c	cgctgaaac	4140	
acccgcac	tctcg	ccatgc	actggccgt	c	cgctgaaac	4200	
atgacatcg	cgtcg	ccatgc	gtagttc	c	cgacttgc	4260	

accggccaggc	gcggtgtggcg	gtgctcgccg	cacacgctcg	cctcgccac	gogcagcgcc	4320
gccgcccagag	cggcgtccgc	ggcgtccgc	aggcgtgcgg	ccaccacgg	aatggcagc	4380
accggtttgt	cctcgaccgt	cgcggccagg	tcgagagcgg	ccagcattag	cacgtatcg	4440
cggtaactggg	ttcgcaatcg	gtgcacgagc	gagccccggca	taccctccga	ttcctcgacg	4500
cactcgacga	acgaccgctg	cagctggta	tgggacggca	gtgtgacctt	gccccgcagc	4560
aatttccagg	actgcggatg	ggcgaccagg	tgatcgccca	acgcccagcga	cgagccccagc	4620
accgagaaca	gccgcccccg	cagactgcgt	tcgcccagca	gagccgcgtt	gagctcggtcc	4680
catccggtgt	ctggatttctc	cgacagccgg	atcaaggcgc	gcagcgcggc	atcgccgtcc	4740
ggagcgcgtg	acagcgacca	cagcaggctcg	acgtgcgcct	gatcctcgat	ccgatccac	4800
cccagctgag	ccagacgcgtc	accagcaggg	gggtcaacta	atccgagccg	gccaacgcgt	4860
ggcaacttcg	gccgctgcgt	ggcgagtttgc	gtcacgacca	cgacggtagc	gcaaagcgcg	4920
tcggcgtcgg	atcaaccgg	agatctggc	tacagcgaca	ggttaggtgcg	cagctcgat	4980
ggcgtgacgt	ggctgcggta	gttcgcccac	tccgtgcgt	tgttgccaa	aaaaaagtca	5040
aaaacgtgct	cccccaaggc	ctccgcacg	agttcggagg	cctccatggc	gcccacgcga	5100
ctatccaaac	tggacggcaa	ttctcggtac	ccatcgctc	ggcgttcctc	gggtgtgagg	5160
tcccatacgt	tgtcctcggc	ctgcggccccc	agcacgtaac	ccttctctac	accccgcaat	5220
cccgccggcca	gcagcacggc	aatgtcaga	tagggattgc	acgcccatac	agggctgcgt	5280
acttcgaccc	gcccgcacga	ggtcttgc	ggcgtgtaca	tcggcacc	cactagggcg	5340
gatcggttgg	cggccccc	cgacgcggcc	gtggcgctt	cggccccc	caccagccgc	5400
ttgttaagagt	tgacccactg	atttgcgacc	gctgtatct	cgcaacgcgt	ctccaggatc	5460
ccggcgtat	acgatattacc	cacttcgac	agctgcagc	gatcatcagc	gctgtggAAC	5520
gcgttgcacat	caccctcgaa	caggctcatg	tgggtgtca	tcggcagcc	cgggtgtgg	5580
ccgaatggct	tgggcatgaa	cgacgcccgg	gcccctt	ccagcgcac	ttctttgtat	5640
acgtacgg	aggcatcac	gttgtcagcc	atcgacagag	cgtccggaaa	ccgcaggatcg	5700
atctcctgct	ggccgggtgc	gccttcgt	tggctgaact	ccaccgagat	gcccacgtat	5760
tccaggccat	cgatcgctg	cgccgcggaa	ttaaggccgg	agtcgtgcac	cgcttggat	5820
aaatagccgg	cgttgtcgac	cgggacgggc	accgaccgt	cctcggtcc	gggcttgcgc	5880
aggaagaact	cgatattcg	atgcacgt	caggagaagc	cgagttcgcc	ggccttcgtc	5940
agctgccg	gcaacacgt	ccgcgggtcc	gcccacgc	gcaacacgt	ccgcgtatgg	6000
atgtcgaaa	acatccgcgc	ttagtgtgg	tggccggaa	tggtgccca	ggcagcacc	6060
tggaaagg	acgggtccgg	gtgcggccacc	gtatcgatt	ccgagaccc	cgcaaagccc	6120
tcgatcgagg	atccgtcgaa	gccgatgcct	tcctcgaa	cgccctcgag	ttcggctgg	6180

gcgatggcga ccgactttag gaaaccgagc acgtctgtga accacagccg gacgaagcgg	6240
atgtcgctt cttccagggt acgaagaacg aattccttct gtccgtccat acctcgaaca	6300
gtatgcactg tctgttaaaa ccgtgttacc gatgcccggc cagaagcggt gcggggcggc	6360
ccgcaagggg agtgcgcggt gagttcaggg cgccgaccgc agactcgctg gccgcgggt	6420
cccgctgaga aaatagtgca tcaccgcaga gtccacacac tggttgcacat cgaacaccgc	6480
agtgtgttgg gtgccgtcga aggtgatcag cggtgccccc agctggcggg ccaggctac	6540
cccgactga tacggagtgg ccgggtcgtg ggtgggtggac accacgacga ccttgcagc	6600
cccgccggc gccgcgggt gcggcgtcga cgttgcggc accggccaca gcgcgcacag	6660
atcgccccgg gcgatccgg tgaactgccc gtagctaagg aacggggcga cctgacggat	6720
cgttggtcg gcggccaccc aggccgctgg atcggccggt gtgggcgcat cgacgcaccc	6780
gaccgcgttg aacgcgtcct ggtcgttgcgt gtagtgcggc tctgcataccc gccgtcata	6840
gtcgtcggca agcaccagca agtcgcggc gtcgtgcgg cgctgcagcc ccagcagacc	6900
actggtcagg tacttccago gctgagggtgt acagcgcgc ttgtatggtgc ccgtcgtcgc	6960
gtcggcgttag ctcaaggccac gtggatccga cgttctaccc ggcttctgca ccagcgggtc	7020
aaccagggcg tggtagcggt tgaccactg ggccgagtcg gtggccagag ggcaggccgg	7080
cgagcgggcg cagtcggcgg cgtagtcatt gaaagcggtc taaaatcccg ccatttggct	7140
gatgcttcc tcgattggc taacggctgg atcgatagcg ccgtcgagga ccatcgcccg	7200
cacatgagta ccgaaccgtt ccaggtaaagc ggtgcccac tcggtgccgt agctgtatcc	7260
gaggtagttg atctgatcgt cacctaacgc ttggcgaacc atgtccatgt cccgtgcac	7320
ggacgcggta ccgatattgg ccaagaagct gaagcccac ccgtcaacac agtcctggc	7380
caactgcccgg tagacctgtt cgacgtgggt gacaccggcc ggactgttagt cggccatcgg	7440
atcgcccccgg tacgcgtcga actcggcgac ggtgcgcacac cgcaacgcag gggtcgagtg	7500
gccgaccctt ctgggtcga agcccaccag gtogaagtgg cggagaatgt cggtgtcgcc	7560
gatcgccgggt gccatagcgg cgaccatgtc gacccggac gccccgggtc cccaggatt	7620
gaccagcagt gctccgaatc gctgtccgt cgccgggacg cggatcaccc ccaacttcgc	7680
ttgtgtccca ccgggttgggt cgtagtcgac gggacggac accgtcgccgc agcgtgcagt	7740
gcgaatttcg ctgggtcgg cgatgaactc gcggcagctg ttccaactct gttgcggcgc	7800
cacgaccggc gcacccgggg ttggccggc gcccgggttct tcagtcgcgc cggccaaacgg	7860
gggcgtgtct agggcagtc cgccgagcag caacccgaag gacagcagcg cggagctcaa	7920
cggtctgcgg cgccacatgg cggccatcgt ctacccggcg aatacctgt acggcgcgaa	7980
atgatcacac ctgcgttct tcgccccgtc agcacttggc gccgcgtggc ggcgtgggtgc	8040
cgccgattaa atacgcgcgtc acgtactcgt caatgcagct gtgcgcctgg aataccaccg	8100

tgtgctgggt tccgtcgaag gtcagcaacg aaccgcgaag ctggttcgcc aggtcgaccc	8160
cggccttgtta cggcgtcgccc gggcatggg tggtgatac caccaccgtc ggcacttaggc	8220
cgggcgccga gacggcatgg ggctgacttg tgggtggcac cgccagaac ggcagggtgc	8280
ccagcggcgc atcaccggtg aacttccgt agctcatgaa cggtgcatc tccgggcgc	8340
ggcggtcttc gtcgatgacc ttgtcgcat cgtaaccgg gggctgatcg acgcaattga	8400
tcgccacccg cgctcaccg gaattttgt agcggccgtg cgagtcccgaa cgcatgtaca	8460
tgtcgccag agccagcagg gtgtctccgc gattgtcgac cagctccgac agccgtcgg	8520
tcaagtgttgc acacagattc ggtgagtaca gcgcataat ggtcccacg atggcgtcgc	8580
tataactcag cccgcgcgga tccttcgtgc gcgcggcct gctgatcctc ggggtgtccg	8640
gtcgaccaa cggatcgacc aggctgttgtt agacctcgac ggctttggcc ggtcgccgc	8700
ccagcgggca gcccgcgttc ttggcgcatcg cggcgccata gttgttgaac gcgtcctgga	8760
agcccttggc ctggcgcage tccgcotcga tggatcgac attgggttcg acggcaccgt	8820
cgagaatcat tgccgcacc cgctgoggaa attcctcgac atacgcggag ccgatccggg	8880
tgcgtacga gtagcccagg taggtcagct tgcgtcgcc caacgcgcg cgaatggcat	8940
ccaggtcctt ggcgacgttg accgtcccgaa catggccag aaagtttttgc cccatcttgt	9000
ccacacagcg accgacgaat tgcttggctcgat ctttcgtgcgt gtcgcacaca ccctccggc	9060
tgttagtcaac ctgcggctcg gcccgcagcc ggtcggtgtc ggcacatcgag ttgcaccaga	9120
tcgccccggg ggacgacgc accccgcggg ggtcgaaccc aaccagggtcg aacctttcg	9180
gcacccgctt cggcaatgtc tggaaagacgc ccaaggcgcc ctgcataaccg gattcgccgg	9240
gtccaccggg atttatgacc agcgaaccga tcttgcgtcc cgtcgcggaa aagcgaatca	9300
gcgcacgcg cggccacgtca ccatcgccggc ggtcgtagtc gaccggtaca gcgagcttgc	9360
cgccataacgc gcccgcgggg atctttactt gccccgttga cgaccggcac ggtgtccact	9420
ccaccggctg gcccagcttc ggctccgcca tacgagcgcc tccccgacc acgcggatgc	9480
agcccacaag aaccaacgc acggcgccga ggcggccca gatcaacacgc atgcgcgcga	9540
tcttgcgcg gcgagacagc ctcatgccc caatgctgcc agagcagacc cgagatctcg	9600
gccagcggcc accgtcgcc gactaaccgg ccgctgcccag cagtcctgcc atcgccgatg	9660
gcgaactcg cggccatccc ccatacgtcc ggtAACAGAT ccggcaaga caccgacccg	9720
tcgaccggat cccgcacggg cgcgtcgcc tcggcggtgc acaactgcga catcagggttgc	9780
gcgcgtggcac cccgtccacg cccgcgttgtt gcacccgttgc catcgccgc gggcgatccc	9840
cgatgcgcgc cacccttcg acgaacccat ctcccacggc ggtcgccggc agcgacgcga	9900
tgtggccgca gatctccgag agttcgccccc gcccggccgg cgacggcaac ccgatgccgt	9960
gcaagtgcacg atcgatgtga ggtcaaggt tcagcgcact gctggcaagc ttttccgaa	10020

accgcggcct cgccttgatc tggagtcaaga accgcgtcacg cagccggtca aaggcgtaac 10080
ccatgctcga gcaaacatgc atgggctgag tggacgttc cagacacagc aactggcgtc 10140
caggccactg agccgctgca tgccgcgtatgg tatgccatgg ggggccccgg gcgcgtctga 10200
gggaaagaag tggcagactg tcagggtccg acgaacccgg ggaccctaac gggccacgag 10260
gatcgaccccg accaccatta gggacagtga tgtctgagca gactatctat ggggccaata 10320
cccccgagg ctccgggccc cgaccaaga tccgcaccca ccacctacag agatggaagg 10380
ccgacggcca caagtgggccc atgctgacgg cctacgacta ttgcacggcc cgatctcg 10440
acgaggccgg catccgggtg ctgctggcgt gtgattcggc gcacaacgtc gtgtacggct 10500
acgacaccac cggtccgatc tccatcgacg agctgatccc gctggtccgt ggcgtggtgc 10560
ggggtgcccc gcacgcactg gtcgtcggc acctgcccgtt cgccagctac gaggcggggc 10620
ccaccggccgc gttggccgccc gccacccggt tcctcaagga cggcggcgca catgcggtca 10680
agctcgaggg cggtgagcgg gtggccgagc aaatcgctg tctgaccgcg ggggcattcc 10740
cggtgatggc acacatcgcc ttccacccgc aaagcgtcaa caccttgggc ggcttccggg 10800
tgcaggccgg cgacgcacgc gccgaacaaa ccatcgccga cgatcgcc gtcgcccgaag 10860
ccggagcggtt tggcgctgtg atggagatgg tgccgcgtt gttggccacc cagatcaccg 10920
gcaagcttac cattccgacg gtcgggatcg gctgtggcc caactgcgac ggccagggtcc 10980
tggtatggca ggacatggcc ggttcagcg gcgcacagac cgccgcgtt gtcaaacgg 11040
atgccgatgt cggtggtgaa ctacgcgtg ctgcaatgca atacgcccgg gaggtggccg 11100
gcggggtatt ccccgctgac gaacacagtt tctgaccaag ccgaatcagc ccgatgcgcg 11160
ggcattgcgg tggcgccctg gatgcgtcg acgcccggatt gccggcgcgg acgcgcacgc 11220
gggaccatc ggcgtcgctg tcgcccgtt agcccggtt gagcccagac attcgatgt 11280
cccaacacca tccgcccacag cccaaatgtat gtggcactct atgcacgcct atccccgacc 11340
aaccaccacc gcggcgacgc atcatgaccg gaggcgaaga tgccagttaga ggcgcggaga 11400
ccagcgcgccc atctggaggt cgagcgcaag ttgcacgtga tcgagtcgac ggtgtcgccg 11460
tcgttcgagg gcatcgccgc ggtgggtcgc gtcgagcgt cgccgaccca gcagctcgac 11520
gcggtgtact tcgacacacc gtcgcacgac ctggcgccga accagatcac cttgcggcgc 11580
cgccacccggc ggcgcgacgc cggctggcat ctgaagctgc cggccggacc cgacaagcgc 11640
accgagatgc gaggcaccgt gtccgcacca ggcgcacgctg tgccggccga gttgttgat 11700
gtgggtctgg cgatcgatcccg cgaccagccg gttcagccgg tcgacggat cagactcac 11760
cgccaaagcc agatcctgtta cggcgccggg ggcgcacgc tggcgaaatt ctgcacgcac 11820
gacgtcaccg catggtcggc cggggcattc cacgcgcgtg gtgcagcggc caacggccct 11880
gccgaacagc agtggcgcga atgggaactg gaactggtca ccacggatgg gaccggcgat 11940

accaagctac tggaccggct agccaaccgg ctgctcgatg ccgggtgccgc acctgccggc 12000
cacggctcca aactggcgcg ggtgctcggt gcgacctctc ccggtgagct gcccaacggc 12060
ccgcagccgc cgccggatcc agtacaccgc gccgtgtccg agcaagtcga gcagctgctg 12120
ctgtgggatc gggccgtcg ggcgcacgcc tatgacgccc tgcaccagat gcgagtgacg 12180
accccaaga tccgcagctt gctgacggat tccaggagt cgtttggcct gaaggaaagt 12240
gcgtgggtca tcgatgaact gcgtgagctg gccgatgtcc tggcgttagc ccgggacgcc 12300
gaggtaactcg gtgaccgcta ccagcgcgaa ctggacgcgc tggcgcggga gctggtaacgc 12360
ggccgggtgc gcgagcgcct ggtagacggg gcgcggcgcc gataccagac cgggctgcgg 12420
cgatcaactga tcgcattgcg gtcgcagcgg tacttccgtc tgctcgacgc tctagacgcg 12480
tttgtgtccg aacgcgcccc tgccacttct ggggaggaat cggcacccgt aaccatcgat 12540
gcggcctacc ggcgagtccg caaagccgca aaagccgcaa agaccgcgg cgaccaggcg 12600
ggcgaccacc acccgacga ggcattgcac ctgatccgca agcgcgcgaa gcgattacgc 12660
tacaccgcgg cggctactgg ggcggacaat gtgtcacaag aagccaaggt catccagacg 12720
ttgctaggcg atcatcaaga cagcgtggtc agccggAAC atctgatcca gcaggccata 12780
gccgcaaca ccgcggcgaa ggacacccctc acctacggtc tgctctacca acaggaagcc 12840
gacttggccg agcgctgcgc ggagcagctt gaagccgcgc tgccgaaact cgacaaggcg 12900
gtccgcaaag cacgggattt agcccgccag ggcggacga gttggctgt aagccggatt 12960
ctgttcccgcc ccccccacagg caagctaacg gccgcacggc ggccgaccatc catctggaca 13020
caccgttacc gggtgcctcg agccgcctac ccgcaggctc gggcgagcaa ccctcaagcg 13080
cctgcgcggc cgcactttcg gtgcggcctt ctggccttgc ttccgggtgg ggttgccta 13140
gccaccccggt tcacccggaa tgctggtgcg ctcttaccgc accgtttcac cttgccacc 13200
acgaggatgg cggctctgttt tctgtggcac ttcccgcga gtcacccctgg attgccgtta 13260
gcaatcaccc tgctctgtga agtccggact ttccctcgact cgacgctgaa cctcgtgaat 13320
ccacacaagc cctacgcgag ccgcggccgc ccagccaaact catccgcgac gaccacgcta 13380
ccccgctggg cgggtgcgcg gccagtgtga ccgcgtggacg acacggctag tcggacagcc 13440
gatccggcgcc gcagtcctta tctgtggactg gtgacacgggt gggacaaacg cgtcgactcc 13500
ggcgactggg acgccatcgcc tgccgagggtc agcgagtcg gtggcgcact gctacccctgg 13560
ctgatcaccc ccggcgaggc cgccggctcg ccgaagctgt acgcccacga cggcctgttt 13620
cgctcgacgg tcgatatggc atccaagcg tacggcgcgc ggcagtatcg atatccat 13680
gccccctatc ccgagtgtatc gagcgtctca agcaggcgct gtatccaaa ctgctgccga 13740
tagcgcgcaa ctggtggccaa actggggcc gggaggcgcc ctggccagac agccttgcgt 13800
actgggtggc gagctgtcat gcccggccaa aacccgatc cacagcgtcg atgttgaagt 13860

acggcaccaa cgactggaac gccctacacc aggtctcta cggcgagtgc gtgttccgc 13920
tgcagggtgt gatcaacctg agcgatccgg aaaccgacta caccggcggc gagttcctgc 13980
ttgtcgaaca gcggcctcgc gcccataccc ggggtaccgc aatgcaactt ccgcaggac 14040
atggttatgt gttcacgacc cgtgatccggc cggtgccggac tagccgtggc tggtcggcat 14100
ctccagtgcg ccatgggctt tcgactattc gttccggcga acgctatgcc atggggctga 14160
tcttcacga cgcagcctga ttgcacgcca tctatagata gcctgtctga ttcaccaatc 14220
gcaccgacga tgccccatcg gcgtagaact cggcgatgct cagcgatgcc agatcaagat 14280
gcaaccgata taggacgccc gacccggcat ccaacgcccag ccgcaacaac attttgc 14340
gcgtgacatg tgacaccacc agcaccgtcg cgccttcgtta gccaacgatg atccgatcac 14400
gtccccgccc aacccggcgc agcacgtcg cgaagcttc cccacccggg ggcgtgatgc 14460
tggtgtcctg cagccagcga cggtgcaact cgggatccgc ttctgcggcc tccgcgaacg 14520
tcagccccctc ccaggcgccc aagtccgtct cgaccaggcgt gtcatcgacg accacgtcca 14580
gggccaggc tctggcggcg gtcaccgcgg tgcgtaaac ccgtgttagc ggccgaggaga 14640
ccaccgcagc gatcccgccc cgccgcgcca gataccggc cgccgcacca acctggcgc 14700
accccacctc gttcaacccc gggttccgc gccccgaata gcggcggtgc tccgacagct 14760
ccgtctgccc gtggcgcaac aaaagtatgc gggtggtgt accgcggcg ccggtccagc 14820
cgggagatgt cggtgactcg gtcgcaacga ttttggcagg atccgatcc gccgcagccg 14880
attgcgcggc ggctccatc gcgtcatgg ccaacccgtc tgcatacgtg ttccggcac 14940
gcggAACCCA ctcgtatgg atcctgcgaa actggacgc caacgcctga gcctggacat 15000
agagcttcag cagatccggg tgcttgaccc tccaccggccc ggacatctgc tccaccacca 15060
gcttggatgc catcagcacc gcggcctcgg tggcacctag tttcacggcg tcgtccaaac 15120
cggttatcag gcccggat ctcgtatgg tgcgtatgg tggcgacgt tggcgatcgc 15180
actcggccag cacggtgag tgatccggg tccacaccac cgccgttat ccggccggc 15240
cgggattgcc ccgcgtatccg ccgtccgtt cgatgacaac tttcactcct caaatcctc 15300
gagccgcaac aagatcgctc cgcatccgg gcagcgcacc acttcatcct ccggccggc 15360
cgagatctgg gccagctcgc cgccggat ctcgtatgg caggcaccac atcgatgacc 15420
ttgcaaccgc ccggccctg gcccgcctcc ggcccgctgt ctggcgatcgc 15480
ctcgatcgtatgg tgcgtatgc gcgtatggcgt gaatgttggt gcccggatcgc 15540
gtcgatccgc gcaagtgcct cgtccaaagc ctgcgtatggcgt ccggccggat cggccgc 15600
cgcttggatgc gcccggat ctcgtatgg tgcgtatgc agtcctcgc ggccgtatcc 15660
cacccatcggc agggcatctt ccaaactggc ttgacggcgt tgcaagctgt cgagctcg 15720
ctgcagatca gccaatttgc tggcgatcgc tgcacccgaa gtgagcaacg accggccggc 15780

gtcgccacgc ttacgcaccc catcgatctc cgactcaaaa cgcgacacct ggccgtccaa 15840
gtcctccgcc gcgattcgca gggccgccc cctgtcgttg gcggcggtgt gctcgccctg 15900
cacctgctgg taagccgccc gctgccccag atgggttagcc cgatgcgcga tccgggtcag 15960
ctcagcatcc agcttcgcca attccagtag cgaccgttgc tgtgccactc cggctttcat 16020
gcctgatctc tcccagtttc gtgatcgagg ttccacgggt cggtgcaagat ggtgcacaca 16080
cgcacccggca gcgacgcgccc gaaatgagac cgcaacactt cggcggcctg gccgcaccac 16140
ggaaattcgc ttgcccataatg cgcgacgtcg atcagggcca cttgcgaagc tcggcaatgc 16200
tcgtcggctg gatgatgtcg cagatcgccc gtaacgtacg cttgcacgtc cggcggcggcc 16260
acggtgccaa gcaacgagtc cccggcggcc cgcagaccg cgacccgcga caccagcagg 16320
tcgggatccc cggcggcgcg cacaccggtc gcagtcggcg gcaacgcggc ctccagacgg 16380
gcaacaaagg tgcgcagcgg ttccgggttt ggcagtctgc caatccggcc taacccgctg 16440
ccgaccggcg gtggtaccag cgcaagatg tcgaatccgg gctcctcgta agggtgcgcg 16500
gcgcgcacatcg cgcacac ctcggcgcgc gtcgtgcgg gtgcgacgac ctcgacccgg 16560
tcctcgccca cccgttcgac ggtaccggac ctgcctatgg cggcgcacgc cccgtcgtgc 16620
gccaggaact gcccggtacc cgacactc cagctgcagt gcgagtagtc gccgatatgg 16680
ccggcaccgg cctcaaagac cgctgcccgc accgcctctg agttctcgcg cggcacatag 16740
atgacccact tgtcgagatc ggccgctccg ggcaccgggt cgagaacggc gtcgacggtc 16800
agaccaacag cgtgtgccag cgctgcggac acacccggcg acgcccggatc ggcgttggtg 16860
tgcgccgtaa acaacgagcg accggccgg atcaggccgt gcaccagcac accctttggc 16920
gtgttggccg cgaccgtatc gacccacgc agtaacaacg ggtggtgcac caatagcagt 16980
ccggcctggg gaacctggtc caccacccgc ggcgtcgct ccaccccaac ggtcaccgaa 17040
tccaccacgt cgtcgggtc gcccacacc agacccacccg aatcccacga ctggcaagc 17100
cgccgggggt aggccctggtc cagcacgtcg atgacatcgg ccagccgcac actcatccgc 17160
gtcctccacg ctttgcac ctcggcgtcg ccgcacccag cacggccac tccggcgcga 17220
ccgcccggcc caggtaccgc gctccaggc cgacgaagggt gtcaccggcg cgcacccgaa 17280
ttcctttgtc ctgcaaatag tttcgtaatc cgtcagcato ggcgtatgg aacagtacga 17340
aaggggccgc accatcgacc acctcgccac ccacccatct cagtcggcc accatctccg 17400
cgccgcacgc cgtcaaccgc accgcacccgg ctgcggcagc ggcgaccgccc cggggggcgc 17460
agcaagcagc gatggccgtc agttgcaatg ttcccaacgg ccagtcgcgt cgctgcacgg 17520
tcaacccgagc cagcacgtct ggcgagccga ggcgttagcc caccgcacccg cggccagcg 17580
accacgtttt cgtcaagcta cggagccca gcacatccgg cagcgagtca tcggccaacg 17640
attgcggctc gccgggaacc caatcagcga acgcctcgac gaccaccagg atgcgtcccg 17700

gccggcgtaa ctcgagcgc tgctcgccga ggtgcagcac cgaggtgggg ttggtcggat 17760
tacccacgac gacaaggctcg gcgtcgctcg gcacgtgcgc ggtgtccagc acgaacggcg 17820
gcttttaggac aacatggtgc gccgtgattc cggcagcgct caaggctatg gccggctcgg 17880
tgaacgcggg cacgacgatt gctgcccga ccggacttag gttgtgcagc aatgcgaatc' 17940
cctccgcccgc cccgacgagc gggagcactt cgtcacgggt tctgccatga cggtcagcga 18000
ccgcgtcttg cgccccgtgc acatcgctgg tgctcgata gcgggcccagc tccggcagca 18060
gcgcggcgcag ctgcccggacc aaccattccg ggggcccgtc atggcggacg ttgacggcga 18120
agtccagcac gccgggcgcg acatcctgat caccgtggta gcgcgcgcgc gcaagcgggc 18180
tagtgtctag actcgccaca gcgtcaaaca gttagtggcc ggtgtgcggg ccaagaatcc 18240
agagcaccgcg cgacgcgttg tctacgcggc gacaaccgcg acatcacagg cagctaacag 18300
ggcgtcggcg gtgatgatcg tcaggccaag cagctgtgcc tgggcgatga gcacacggtc 18360
gaatggatgt cgatggtgat ccggaagctc tgccgtgcgc agtgtgtgcg tggtaactg 18420
acagcggcga cgtgccgcag cggcgcattc gatcgggcac gtaagaagcc gatggctcgg 18480
gcggcgggag cttgccgagg cggtagttga tcgcgcattc ccaggcactg gcggccgaca 18540
agagaatgct gttgcggacg tcctgaacaa tcgcccgtgt ttgcgttgcg gcatccgcag 18600
ccaaacgtgg gtgtcgatga ggtagcgctt caccggtaa agcgttcgag cacgtcgtct 18660
gacaacggag cgtccaaatc gtcgggcacg cgttacacgc catggtaat gcctaaccgc 18720
cgagtcctcat gaggatgcag cggcacaaggc tttgctaccg gtcgcgcgcg gcgggcaatc 18780
tcaacctctg cccgcgttag acgagccgca gcagctcgga caggcgtgtc ttgcgcctcg 18840
gaacgcgcac cgcgttcgca ggcgcggcaga cttcgcgtc gaccacctgc tcaccaaact 18900
tcgcgcattcat cgcctgatcat cacagcgcca acgggttagcg gtttgcgttccaa ccgcgttcgtc 18960
aacgacaatg ggatcggtac cgacacgacc gcgagcggga ccaattgcgc gcctcctcca 19020
cgcgccgccc cacggcgccgc atcgctcgccg ggtgaatcgc cgacgttgt gatcttcgt 19080
ctggacggca cgctgaccga ctcggcgccgc ggaatcgat ccagcttccg acacgcgcgc 19140
aaccacatcg gtgccccagt acccgaaggc gacctggcca ctcacatcg tggcccgccc 19200
atgcgtgaga cgctgcgcgc catggggctc ggcgaatccg ccgaggaggc gatcgtagcc 19260
taccggggccg actacagcgc cgcgggttgg gcatgtgaaca gttgttcga cggatcgcc 19320
ccgctgtgg ccgacctgctcg caccggcggt gtccggctgg ccgtcgccac ctccaaaggca 19380
gagccgcaccg cacggcgaaat cctgcgcac ttcggatttgc agcagcactt cgaggtcattc 19440
gcggcgcgcgac gacccgttgg ctcgcgaggc agcaaggctcg acgtgttgc ccacgcgcgc 19500
gcgcagctgc ggccgctacc cgagcggttgc ggtatggtgc ggcggccag ccacgcgcgc 19560
gacggggccgg cgcgcacgg catcgacacg gtgggtggtgc gctggggctaa cgggcgcgc 19620

gactttatcg acaagacctc caccaccgtc gtgacgcatttccgacatg ccggccacatg tgacgagctg 19680
agggaggcgc taggtgtctg atccgctgca cgtcacatttccgacatg gcaacatctg 19740
ccgggtcgcca atggccgaga agatgttcgc ccaacagctt cgccaccgtg gcctgggtga 19800
cgcggtgcga gtgaccagtgcg cgggcaccgg gaactggcat gtaggcagtt gcggccacgaa 19860
gcggggcggcc ggggtgttgc gagcccacgg ctaccctacc gaccaccggg ccgcacaagt 19920
cggcaccgaa cacctggcgg cagacctgtt ggtggcatttgc gaccgcaacc acgctcggt 19980
gttgcggcag ctcggcgtcg aagccgccccg ggtacggatg ctgcggcat tcgacccacg 20040
ctcgaaaacc catgcgctcg atgtcgagga tccctactat ggcatcact ccgacttcga 20100
ggaggtcttc gccgtcatcg aatccgcctt gcccggcctg cacgactggg tcgacgaacg 20160
tctcgcgcgg aacggaccga gttgatgccc cgcctagcgt tcctgctgcg gcccggctgg 20220
ctggcggttgc ccctggcgtgttgc acctacctgt gctttacggt gctcgcgcgg 20280
tggcagctgg gcaagaatgc caaaacgtca cgagagaacc agcagatcag gtattccctc 20340
gacaccccccgc cggttccgct gaaaaccctt ctaccacagc aggatttgcg ggcgcggac 20400
gcccgcgtggc gcccgggtgac ggcaaccggc cagttacccgc cggacgtgca ggtgctggcc 20460
cgactgcgcg tggtgaggg ggaccaggcg tttgaggtgt tggcccccatt cgtggcgtac 20520
ggcggaccaa ccgttctggt cgaccgtgga tacgtgcggc cccaggtggg ctgcacgtta 20580
ccaccgatcc cccgcctgccc ggtgcagacg gtgaccatca ccgcgcggct gcgtgactcc 20640
gaaccgagcg tggcgggcaaa agaccatttgc gtcagagacg gttccagca ggtgtattcg 20700
atcaataccg gacaggtcgc cgcgtgcacc ggagtccagc tggctgggtc ctatctgcag 20760
ttgatcgaag accaaccggc cgggctcgcc gtgctcgccg ttccgcatttgc agatccggg 20820
ccgttctgttgc cctatggcat ccaatggatc tggctggca ttctggcacc gatcggtttg 20880
ggctatttgc cctacgcccggatccggccg cccggccgggg aaaaaggggg gtcgcccacca 20940
ccggacaaggc caatgacggt cgagcagaaaa ctcgtgcacc gctacggccg ccggcggtaa 21000
accaacatca cggccaataac cgcagccccccgc gcttggacca cccgcgcacag caccacggcg 21060
cgccgcagat cggccacccgtt gggcgaccgg ccgtcgccca aggtggccg gatctgcac 21120
tcgtggcgttgc accgggtggg cccaccaggc cgcacgtcaa ggcggccacg aaacggccccc 21180
tcgtggcgttgc accgggtggg gctggatgg cggcgccgt cgccgcggccca ggcccgtacc 21240
gcaccggccgg gcgacccacc gaccaccggc ggcgcagatca ccaccagcac cgccgtcgcc 21300
cgtgcgcggc catagttggc ccagtcatcc aatcggtctg cagcccaacc gaatcgagaa 21360
taacgcccggc agcggtagcc gatcatcgag tccagggtgt tgatggcacc atatccggc 21420
accgcaggca cggccgtcgaa agccgcccac agcagcggca ccacccggc gtcggccgg 21480
ttttcgccca ccgactccag cgcggcacgc gtcaggcccg ggccgcccacg ctggccgggg 21540

tcacccccgc acagcgacgg cagcagccgt cgccgcgcct cgacatogtc gcgctccaac 21600
aggccgata tctggcggcc ggtgcgcgcc agcgaagtgc cgcccagcgc tgcccaggta 21660
gccgtcgccgg tggccgcccac gggccaggac ctgcccggta gccgctgcag tgccgcgcg 21720
agcaagccca ccgcgcgcgac cagcaggccg acgtgtaccg caccggcgac ccggccgtca 21780
cggttaggtga tctgctccag cttggcggcc gcccgaccga acagggccac cggatgaccc 21840
cgtttgggtt cgccgaacac gacgtcgagc aggccagccga tcagcacgac gacggccctg 21900
gtctgccagg tcgatgcaaa cactccggca gcgtcgacaca cgtggtctac gctcagctat 21960
ttatgacctc atacggcagc tatccacgat gaagcggcca gctacccggg ttggcgcacct 22020
gttgaaccccg gcggcaatgt ttttgccggc agcgaatgtc atcatgcagc tggcagtgcc 22080
gggtgtcggg tatggcgtgc tggaaagccc ggtggacagc ggcaacgtct acaagcatcc 22140
gttcaagccgg gcccggacca ccggcaccta cctggcggtg gcgaccatcg ggacggaatc 22200
cgaccgagcg ctgatccggg gtggcgtggc cgtcgccac cggcagggttc ggtcgacggc 22260
ctcgagccca gtgtcctata acgccttcga cccgaagttt cagctgtggg tggcggcgtg 22320
tctgtaccgc tacttcgtgg accagcacga gtttctgtac ggcccactcg aagatgccac 22380
cgccgacgcc gtctaccaag acgccaaacg gttagggacc acgctgcagg tgccggaggg 22440
gatgtggccg ccggaccggg tcgcgttcga cgagtactgg aagcgctcgc ttgatggct 22500
gcagatcgac gcgcgggtgc gcgagcatct tcgcgggggtg gcctcggtag cgtttctccc 22560
gtggccgttg cgcgccgtgg ccggggcggtt caacctgttt gcgacgacgg gattcttggc 22620
accggagttc cgcgcgatga tgcagctggc gtggtcacag gcccgacgc gtcgcitcga 22680
gtggttactt tccgtctac gtttagccga cccgctgatt ccgcacatggg cctggatctt 22740
cgtttaccag cttaacttgtt gggacatgcg gtttcgcgcgc cgacacggcc gccgaatcgt 22800
ctgatagagc ccggcccgagt gtgagcctga cagcccgacca ccggcggcgt gtgtcgccgc 22860
gccaggttca cgctcgccga tctagagccg ccgaaaacct acttctgggt tgcctcccg 22920
atcaacgtgc tgcgtctac gagcagctca cgcataatcg cgcgcacgc atccacccgc 22980
gcatacaggt ccgccttggc cgccggcagc tggccgcacg tcattggccg caccggcgt 23040
gctgtctgtc gcgcgcgcgt gtcgtttga aacccaggta gtcacccac gaccacgaca 23100
ctgccccatcc cggcgcccccg ccgacaacgac agcacagcta gccgggtggc gggacggga 23160
tcgaaccggcc gaccgctggc gtgtaaaacc agagctctac cgctgagcta cgcccccatt 23220
accggcccgag gctacacgc ttgcggccaa gcacccaaaa ccttaggccc taagcgccgc 23280
cagagcgctcg gtccacagcc gctgatcgac aacttcaccc ggctgcttca tctcggcga 23340
ccgaatgatc cctgaccgat cgaccacaaa ggtgccccgg ttagcgatgc cgccctgctc 23400
gttgaagacg ccgtaggcct gactgaccgc gccgtgtggc cagaagtccg acaacagcgg 23460

aaacgtgaat ccgctctgcg tcgcccagat ctgtgagt ggtggcggc ccaccgaaat 23520
cgctagcgcg gcgcgtcgct cggtctcaaa ctcggcagg tgatcacgca actggtccag 23580
ctgcctgg cagatgccc tgaacgcca cgaaaagaac accaacagca cgttcttgc 23640
accccgtag cccgcgcagg tgacaagctg ctgattctgg tcgcgcacg tgaagtcagg 23700
ggcggtggtc cccacgttca gcatcagcgc ttgcgcaggc gcgatttcgg ctgtaccaat 23760
ctgctggcgc tccagttgcc cagattgacc gacgaggctg gcatcagccc agctgtggc 23820
gccgcctcg caatctcgcc gggcaataca tgccgggtc ggccggtctt gggcgtcacc 23880
acccaaatca caccgtcctc ggccgcggg ccgatcgcat ccatcagggt gtccaccaaa 23940
tcgcccgtcgc catcacgcca ccacaacagg acgacatcga tgacctcgic ggtgtcttca 24000
tcgagcaact ctcccccgca cgcttcttcg atggccgcgc ggatgtcgic gtcggtgtct 24060
tcgtcccagc cccattcctg gataagttgg tctcggttga tgcccaattt gcggcgttag 24120
ttcgaggcgt gatccgcgc gaccaccgtg gAACCTCCTT cagtctcgic gggccatgtg 24180
cacaccgtcg cgatggcgt tatcgicgc cagccagaac cggccaccc gcccgcctca 24240
gaaggcggcc acgcacattt tcaatgcctt tgtcttggtg tcgttgcgc gatcaacccg 24300
ccgggtgaat tccgctgtcg acgcgtgcgc accgatggca tttgccaccg cgccggccgc 24360
gtcgacatat gcgttgcgc catccccag ttgcgcggac agcgcgcgc tcagactgcc 24420
tgagaccgtc gaggcactgt ttttgcgcgc gtcgatggcc ggacccctgg tcggcccggt 24480
gttgcggccc tgattgaacg cggccacgta ggcgttcacc ttgtcgatgg cgtccttgc 24540
ggtggccgcg acgcgtcgc acgagggtcg aatcgccctt gtcgtcagcg attgttggcg 24600
ctgcgactcc cggatgtcg acgtcgccgc cgaagccgac accgacgcgg acacccgacga 24660
gcggtaggcc ggtgcgcacgt tggtgtcgcatggccatggccatggc cagttggata 24720
tccgacgatc cccatcagca gcagcgcat gcagccgagc gccaggcgcctcgc 24780
gagctcccccc cctgtgcctgc gaggcacggc ggcgcacccg atgagcacgg catgtgaggt 24840
tacctggtcg cagcgccgacc gcgcgtggcc tggtgtgtcg cgcacccgcga gaaccgagcg 24900
gagtgcggct atccgcccgc gacgcgggtg cggcacgata gggggacgac catctaaaca 24960
gcacgcacgc ggaagccgcg cacctacagg agtagtgcgt tgaccaccga ttgcggccgc 25020
cacatctgg cccaaactc aaacagcgca agcgaacccg accgaggctcg ggtgtccgc 25080
gagggtgtgg cgtcgatattt gcccgcacattt gatccccagg agaccccgta gtggctggag 25140
tccttgcaca cgctgcgtca acgcgtcgccgc cgcgtcgccggg cccgcgtaccc gatgttgcgg 25200
ctgctagagc gggccggcga gcagcgggtg gccatcccg cattgacgtc taccgactat 25260
gtcaacacca tcccgaccga gctggagccg tgggtccccc ggcacgaaaga cgtcgaacgt 25320
cgttatcgag cgtggatcag atggaatgcg gccatcatgg tgcaccgtgc gcaacgaccg 25380

ggtgtggcg tgggtggcca tatctgacc tacgcgtcgt ccgcggcgct ctatgaggtc 25440
ggttcaacc acttcttccg cgcaagtcg caccggcg gcccgcataa ggtgttcata 25500
cagggccacg ctccccggg aatctacgcg cgcccttcc tcgaaggcg gttgaccgac 25560
gagcaactcg acggattccg ccaggaacac agccatgtcg gcggcgggtt gccgtcctat 25620
ccgcacccgc ggctcatgcc cgacttctgg gaattccca ccgtgtcgat gggtttggc 25680
ccgctcaacg ccatctacca ggcacggttc aaccactatc tgcatgaccg cggtatcaaa 25740
gacacctccg atcaacacgt gtggtgaaaa ttgggcgacg gcgagatgga cgaacccgag 25800
agccgtggc tggcccacgt cgccgcgtg gaaggcttgg acaacttgac cttcgtgatc 25860
aactgcaatc tgcagcgact cgacggcccg gtgcgcggca acggcaagat catccaggag 25920
ctggagtcgt tcttccgcgg tgccggctgg aacgtcatca aggtggtgcg gggccgcgaa 25980
tggatgccc tgctgcacgc cgaccgcgac ggtgcgtgg tgaatttaat gaataacaaca 26040
cccgatggcg attaccagac ctataaggcc aacgacggcg gctacgtcg tgaccacttc 26100
ttcggccgcg acccacgcac caaggcgctg gtggagaaca tgagcgacca ggatatctgg 26160
aacctcaaac gggcgccca cgattaccgc aaggtttacg ccgccttaccg cgccgcgc 26220
gaccacaagg gacagccgac ggtgatccgc gccaagacca tcaaaggcta cgcgctggc 26280
aagcatttcg aaggacgca tgccacccac cagatgaaaa aactgaccct ggaagacctt 26340
aaggagtttc gtgacacgcgca goggattccg gtacgcgacg cccagcttga agagaatccg 26400
tacctgcgc cctactacca ccccgccctc aacgccccgg agattcgta catgctcgac 26460
cgccgcggg ccctcggggg ctgtttccc gagcgcagga ccaagtccaa agcgctgacc 26520
ctgccccgtc ggcacatcta cgccgcgtg aaaaagggtt ctgggcacca ggaggtggcc 26580
accaccatgg cgacgggtcg cacgttcaaa gaagtgtgc ggcacaaacca gatcgcccg 26640
cgatggatgtcc cgatcattcc cgacgaggcc cgacacccgt ggatggactc ctggttcccg 26700
tcgctaaaga tctataaccg caatggccag ctgtataccg cggttgacgc cgacctgatg 26760
ctggcctaca aggagagcga agtggggcag atcctgcacg agggcatcaa cgaagccggg 26820
tcgggtggct cgttcatcgcc ggccggcacc tcgtatgcga cgcacaaacga accgatgatc 26880
cccatttaca tcttctactc gatgttccgc ttccagcgca cccgcgatag ctctggcc 26940
gcggccgacc agatggctcg agggttcgat ctcggggcca cccgcggcg caccaccctg 27000
accggtgagg gcctgcaaca cccgcacggt cactcgatc tgctggccgc caccaccccg 27060
gcgggtgggt cctacgaccc ggccctcgcc tacgaaatcg cctacatcg ggaaagcgga 27120
ctggccagga tgtgcggggaa aacatcttct tctacatcac cgtctacaac 27180
gagccgtacg tgcagccgac ggagccggag aacttcgatc ccgaggcggt gctgcgggg 27240
atctaccgt atcacgcggc caccgagcaa cgcaccaaca aggcgcagat cctggcctcc 27300

gggttagcga tgcccgcggc gctgcgggca gcacagatgc tggccgccga gtggatgtc 27360
gccgcccacg tgtggtcggt gaccagttgg ggcgagctaa accgcgcacgg ggtggccatc 27420
gagaccgaga agctccgcca cccccatcg cgccgcggcg tgccctacgt gacgagagcg 27480
ctggagaatg ctggggcccg ggtgatcgcg gtgtcgact ggatgocgcg ggtccccgag 27540
cagatccgac cgtgggtgcc gggcacatac ctcacgttgg gcacccgacgg gttcggctt 27600
tccgacactc ggcccgcgc tcgcccgtac ttcaacaccg acgcccgaatc ccagggtggc 27660
gcggttttgg aggcggttggc gggcgacggc gagatcgacc catcggtgcc ggtcgccggc 27720
gcccgcagt accggatcga cgacgtggcg gctgcgccc gacgacccac ggatcccgg 27780
cccgccccct aacgcccggcg agccgaccgc ctttggccga atcttccaga aatctggcgt 27840
agcttttagg agtgaacgac aatcagttgg ctccagttgc cccggcagg tcggcgctcg 27900
aactgctgga cactgtgccc gattcgctgc tgccgcgggtt gaagcagtac tcggccggc 27960
tggccaccga ggcagttcg gccatgcaag aacggttgcc gttttcgcc gacctagaag 28020
cgtcccaagcg cgccagcgtg ggcgtggcg tgcagacggc cgtggtaaac ttctcgaaat 28080
ggatgcacga cccgcacagt gacgtcggtc ataccgcga ggcattcgag ctggtcccc 28140
aggatctgac ggcacggatc ggcgtcgcc agaccgtgga catggtcgg gtcaccatgg 28200
agttttcga agaagtcgtg cccctgctcg cccgttccga agacgagttt accgcctca 28260
cggtggcat tttgaaatac agccgcgacc tggcattcac cgccgcacg gcctacgccc 28320
atgcggccga ggcacggacc acctggaca gccggatgga ggccagcgtg gtggacgcgg 28380
tggtacgccgg cgacaccgggt cccgagctgc tgtcccgccg ggccgcgtg aattgggaca 28440
ccacogcgcc ggcgaccgtt ctggggaa ctccggcgcc cggtaaaat ggctccaaca 28500
gcgacggcga cagcgagcgg gccagccagg atgtccgcga caccgcggct cgccacggcc 28560
gcgctcgct gaccgacgtg cacggcacct ggctggtgcc gatcgctcc ggccagctgt 28620
cgccaaaccga gaagttcctc aaagacctgc tggcagcatt cgccgacgccc ccggtggtca 28680
tcggccccac ggcgcccattt ctgaccgcgg cgccacgcag cgctagcggag gcatctccg 28740
ggatgaacgc cgtcgcccggc tggcgcggag cgccgcggcc cgtgtggct agggacttt 28800
tgccogaacg cgccctgatg ggcgacgcct cggcgatcgt ggccctgcat accgacgtga 28860
tgcggccccct agccgatgcc ggaccgacgc tcacgcgac gctagacgc tatctggatt 28920
gtggccgcgc gattgaagct tgtgcagaa agttgttcgt tcacccaaac acagtgcgg 28980
accggctcaa gcggatcacc gacttcacccg ggcgcgatcc caccgcgca cgccatgcct 29040
atgtccttcg ggtggccggcc accgtgggtc aactcaacta tccgacgcgc cactgaagca 29100
tcgacagcaa tgccgtgtca tagattccct cggccgtcag aggggtcca gcagggggccc 29160
cgaaaagata ccaggggcgc cgtcggacgg aaagtgtatcc agacaacagg tcgcgggacg 29220

atctcaaaaa catagttac aggcccgtt tgggttat atacaaaaac ctaagacgag 29280
 gttcataatc tggtacaccg cgcaaaaccg tcttcacagt gttctcttag acacgtgatt 29340
 gcgttgctcg cacccggaca gggttcgcaa accgaggaa tgggtcgcc gtggcttcag 29400
 ctgcccggcg cagcggacca gatcgcggcg tggcgaaag ccgctgatct agatcttgcc 29460
 cggctggca ccaccgcctc gaccgaggag atcaccgaca ccgcggtcgc ccagccattg 29520
 atcgctcgccg cgactctgct ggcccaccag gaactggcgc gccgatgcgt gctcgccggc 29580
 aaggacgtca tcgtggccgg ccactccgtc ggcgaaatcg cggcctacgc aatcgccggt 29640
 gtgatagccg ccgacgacgc cgtcgcgctg gccgccaccc gcggcgccga gatggccaag 29700
 gcctgcgcca ccgagccgac cggcatgtct gcgggtcgctg gcggcgacga gaccgagggt 29760
 ctgagtcgcc tcgagcagct cgacttggtc cccgcaaaccc gcaacgcgcg cggccagatc 29820
 gtcgctgccc gccggctgac cgcgttggag aagctcgccg aagacccgccc ggccaaggcg 29880
 cgggtcgctg caactgggtgt cggcggagcg ttccacaccg agttcatggc gcccgcactt 29940
 gacggctttg cggcggccgc ggccaacatc gcaaccgccc accccaccgc cagctgctg 30000
 tccaaccgcg acgggaagcc ggtgacatcc gccggccgcgg cgatggacac cctggctcc 30060
 cagctcaccc aaccgggtcg atgggacctg tgcaccgcga cgctgcgcga acacacagtc 30120
 acggcgatcg tggagttccc ccccgccggc acgcttagcg gtatcgccaa acgcaactt 30180
 cgggggggttc cggcacgcgc cgtcaagtca cccgcagacc tggacgagct ggcaaaccta 30240
 taaccgcgga ctggccaga acaaccacat acccgtcaat tcgatttgta cacaacatat 30300
 tacgaaggga agcatgctgt gcctgtcaact caggaagaaa tcattggccg tatcgccgag 30360
 atcatcgaag aggttaaccgg tatcgagccg tccgagatca ccccgagaa gtcgttcgtc 30420
 gacgacctgg acatcgactc gctgtcgatg gtcgagatcg ccgtgcagac cgaggacaag 30480
 tacggcgtca agatccccga cgaggacctc gccggctctgc gtaccgtcg tgacgttgc 30540
 gcctacatcc agaagctcgaa ggaagaaaac ccggaggccg ctcaggcggt ggcgcgcaag 30600
 attgagtcgg agaaccggc tgccgttgcc aacgttcagg cgaggcttga ggccgagtcc 30660
 aagtgagtca gccttccacc gctaattggcg gtttccccag cggtgtgggt accggcgatca 30720
 cagcgacgac gtcgatctcg ccggacatcg agagcacgtg gaagggtctg ttggccggcg 30780
 agagcggcat ccacgcactc gaagacgagt tcgtcaccaa gtgggatcta ggggtcaaga 30840
 tcggcgggtca cctcaaggat ccgggtcgaca gccacatggg ccgactcgac atgcgacgca 30900
 tgtcgtacgt ccagcggatg ggcaagttgc tggcgccgaca gctatggag tccggccggca 30960
 gcccggaggt cgatccagac cggttcgcgc ttgttgcgg caccggctta ggtggagccg 31020
 agaggattgt cgagagctac gacctgatga atgcggccgg ccccccggaaag gtgtccccgc 31080
 tggccgttca gatgatcatg cccaaacggtg ccgcggccgt gatcggtctg cagttgggg 31140

cccgccgcgg ggtgatgacc cccgtgtcgg cctgttcgtc gggctcgaa gcgatcgccc 31200
acgcgtggcg tcagatcgtg atgggcgacg ccgacgtcgc cgtctcgcc ggtgtcgaag 31260
gaccatcga ggctgtcccc atcgccgcgt tctccatgtat gcggggccatg tcgaccgc 31320
acgacgagcc tgagcgggcc tcccgccgt tcgacaagga ccgcgacggc tttgttgc 31380
gcgaggccgg tgcgtgatg ctcatcgaga cgaggagca cgccaaagcc cggtggccca 31440
agccgttggc ccgattgctg ggtgccggta tcacctcgga cgccttcata atggtggcgc 31500
ccgcggccga tggtgttcgt gccggtaggg cgatgactcg ctgcgtggag ctggccgggt 31560
tgtcgccggc ggacatcgac cacgtcaacg cgacggcac ggcgacgcct atcgccgacg 31620
ccgcggaggc caacgcccata cgctcgccg gtgtgatca ggccgcggtg tacgcgcga 31680
agtctgcgtt gggccactcg atcgccgcgg tcgggtcgct cgagtcggc ttcacggcgc 31740
tgacgctgcg cgacggcgct atccgcga ccctgaacta cgagacaccc gatcccgaga 31800
tcgaccttga cgtcgccgc ggcgaaccgc gctatggcga ttaccgtac gcagtcaaca 31860
actcggttgcg gttcgccggc cacaatgtgg cgcttgcctt cggcgttac tgaagcacga 31920
catcgccgggt cgcgaggccc gaggtggggg tccccccgt tgccggggcg agtcggaccg 31980
atatgaaagg aacgttcgca agaccaatga cggagctgg taccggaaa gcctttccct 32040
acgttagtcgt caccggcata cccatgacga ccgcgtcgac gaccgacgcg gagactacgt 32100
ggaagttgtt gctggaccgc caaagcggga tccgtacgct cgatgaccca ttgcgtcgagg 32160
agttcgacct gccagttcgc atcgccggac atctgcttga ggaattcgac caccagctga 32220
cgccgatcga actgcgcggg atggataacc tgcagcggat gtccaccgtg ctgagccggc 32280
gcctgtggga aaatgcggc tcacccgagg tggacaccaa tcgattgtatg gtgtccatcg 32340
gcaccggcct gggttcgcc gaggaactgg tcttcagttt cgcgcgtatg cgctcgccg 32400
gaatgaaggc ggtctcgccg ctgaccgtgc agaagtacat gccaacggg gcccggccgg 32460
cggtcgggtt ggaacggcac gccaaggccg ggggtatgac gcccgtatcg gcgtgcgcata 32520
ccggccgcga gcccattcgcc cgtcgccggc agcagattgt gctggagag gccgatgccc 32580
ccatctgcgg cggcgtggag accaggatcg aagcggtgcc catgcgcggg ttgcgtcaga 32640
tgcgcacatcgatgtccacc aacaacgacg accccgcggg tgcatgcgc ccattcgaca 32700
gggaccgcga cggcttgcgtt ttcggcagg gcccgcctt tctgttgcata gagaccgagg 32760
agcacgcac ggcacgtggc gccaacatcc tggccggat catggcgcc agcatcacct 32820
ccgatggctt ccacatggtg gccccggacc ccaacggggc acgcgcggg catgcgtt 32880
cgccggcgat tcagctggcg ggcctcgccc ccggcgacat cgaccacgtc aatgcgcacg 32940
ccaccggcac ccaggtcgcc gacctggccg aaggcagggc catcaacaac gccttggcc 33000
gcaaccgacc ggcgggttac gcccccaagt ctgcctcggtt ccactcggtg ggcgcggcgt 33060

gcgccgtcga atcgatcttg acgggtctcg cgttgcgca tcaggtgatc ccggccgacac 33120
 tgaatctggt aaacctcgat cccgagatcg atttggacgt ggtggcggtt gaaccgcac 33180
 cgggcaatta ccggtatgcg atcaataact cgttcggatt cggcggccac aacgtggcaa 33240
 tcgccttcgg acggtaactaa accccagcgt tacgcgacag gagacctgcg atgacaatca 33300
 tggcccccgaa ggcgggtggc gagtcgctcg acccccgca tccgctgttgc cggtgagca 33360
 acttcttcga cgacggcgcg gtggaaattgc tgacgcgacg tgaccgctcc ggagtgcgtgg 33420
 ccgcggcggg caccgtcaac ggtgtgcgca ccacgcgtt ctgcaccgac ggcaccgtga 33480
 tggcggcgc catggcggtc gaggggtgca cgcacatcgta caacgcctac gacactgcca 33540
 tcgaagacca gagtcccattt gtgggcattt ggcattcggg tggtgcccg ctggctgaag 33600
 gtgtgcgggc gctgcacgcg gttaggcagg tggtaaaggc catgatccgc gctccggct 33660
 acatccccca gatctcggtt gtcgtcggtt tggccggcccg cggcggccccc tacggaccgg 33720
 cgttgaccga cgtcgctcgat atggcgccgg aaagccgggt gttcgtaacc gggcccgacg 33780
 tggtgccgcg cgtcaccggc gaggacgtcg acatggcctc gtcgggtggg ccggagaccc 33840
 accacaagaa gtccgggggtt tgccacatcg tggccgacga cgaactcgat gcctacgacc 33900
 gtggcgccg gttggtcgga ttgttctgcc agcagggca ttgcgtcgatc agcaaggccg 33960
 aggccggta caccgacatc cacgcgtc tggccgaaatc ctcgcgacgt gcctacgacg 34020
 tgcgtccgat cgtgacggcg atcctcgatg cggacacacc gttcgacgag ttccaggcca 34080
 attggcgcc gtcgatgggtt gtcgggttgg gtcgggtgtc gggtcgcacg gtgggtgtac 34140
 tggccaaaca cccgctacgc ctggcgccgt gcctgaactc cgaaagcgca gagaaggcag 34200
 cgcgttctgtt gcccgtgtgc gacgcgttgc ggattccgct ggtgggtggt gtcgatgtgc 34260
 cgggctatct gcccgtgtc gaccaggagt ggggtggcgt ggtgcggccgt ggcgccaagt 34320
 tgctgcacgc gttcggtcgag tgcaccgttc cgcgggtcac gtcgggtcacc cgaaagaccc 34380
 acggcggggc atacattgcg atgaactccc ggtcggtgaa cgcgaccaag gtgttcgcct 34440
 ggccggacgc cgagggtcgcg gtgatggcg ctaaggccgc cgtcggcatc ctgcacaaga 34500
 agaagttggc cgccgctccg gagcacgaac gcaagcgct gcacgaccag ttggccgccc 34560
 agcatgagcg catcgccggc ggggtcgaca gtgcgttggc catcggtgtt gtcgacgaga 34620
 agatcgaccc ggccgataact cgcagcaagc tcaccgaggc gtcggcgacg gtcggccac 34680
 ggcggcccg ccacaagaac atcccgctgt agttctgacc gcaagcagac gcagaatcg 34740
 acgcgcgagg tccgcggccgt gcgattctgc gtctgcgtc cagttatccc cagcggtggc 34800
 tggtaaacgc gaggcgctcc tgcgtcgatc ggacggtgcc taccgacgcg ctaacaattc 34860
 tcgagaaggc cggcgggttc gccaccaccg cgcattgtc caccgtcatg acccgccaac 34920
 agtcgacgt ccaagtgaaa aacggcggcc tcgatcgatc ttggtaacggg gtctacgcgg 34980

cacaagagcc ggacctgttg ggccgcttgg cggcttcga tgtgttcatg ggggggcacg	35040
ccgtcgctg tctgggcacc gcccggcgt tgtatggatt cgacacggaa aacaccgtcg	35100
ctatccatat gctcgatccc ggagtaagga tgcggccac ggtcggtctg atggtccacc	35160
aacgcgtcgg tgcccggtc caacgggtgt caggtcgctc cgccgaccgcg cccgcatgga	35220
ctgccgtgga ggtcgcacga cagttgcgcc gcccgcggc gctggccacc ctgcacgccc	35280
cactacggtc aatgegctgc gctcgcaagt aaattgaaaa cgccgttgct gagcagcggag	35340
gcccggagg catcgtcgcg gcgcgcgaac tcttaccctt cgccgacgga cgccggaat	35400
cggccatgga gagcgaggct cggctcgta tgatcgacca cgggctgccc ttgcccgaac	35460
ttcaatacccc gatacacggc cacggtggtg aaatgtggcg agtgcacttc gcctggccc	35520
acatgcgtct cgccggcga tacgaaagca tcgagtggca cgcgggaccg gcggagatgc	35580
tgcgcgacaa gacacgctgg gccaagctcc aagagctcggtt gttggacgatt gtcccgattt	35640
tcgtcgacga tgtcagacgc gaaccggcc gcctggccgc ccgcacatcgcc cgccacctcg	35700
accgcgcgcg tatggccggc tgaccgctgg tgagcagacg cagagtcgca ctgcggccgg	35760
cgcagtgcga ctctgcgtct gctcgctc aacggctgag gaactcctta gccacggcga	35820
ctacgcgtc gcgatcccggtt ggcaccagac cgatccgggtt ccggcggtcg aggatatcg	35880
ccacatccag cgccccctca tgggtcacccg cgtattcgaa ctccgcccgg gtcacgtcga	35940
tgccgtcgcc gaccggctcg gtggggccctt cacatgtggc ggccgcagcg acgttggccg	36000
cctcgccccc gtaccgcgcc accagcgact cgggcaatcc ggcccccgtat ccggggcccg	36060
gcccagggtt cgccgggtcg ccgatcagcg gcaggttgcg agtgcggcac ttgcggctc	36120
gcaggtgtcg cagcgtgatg gcgcgatitca gcacatcctc tgccatgttag cggtattccg	36180
ttagcttgcc gccgaccaca ctgatcacgc ccgacggcga ttcaaaaaca gcgtggtcac	36240
gcgaaacgtc ggccgtgcgg ccctggacac cagcaccggcc ggtgtcgatt agcggccgca	36300
atccccata ggcaccgtatg acatccttgg tgccgaccgc cgtccccaat gcggtgttca	36360
ccgttatccag caggaacgtg atcttcccg aagacggttt gggcacatcg ggaatcgccg	36420
cggtgcgtc ttgcgtcggtc agcccgagat agatccggcc cagctgctcg ggcatggcga	36480
acacgaagcg gttcagctca ccggggatcg gaatggtcag cggccagtc ggattggcaa	36540
acgacttcgc gtcgaagacc agatgtgtgc cggcgctgg gcgtagcctc agggacgggt	36600
cgatctcacc cgcacacacg cccggccgt tgatgacggc acgcgcgcac agcgcgaacg	36660
actgcccgggt gcgcgggtcg gtcaactcca ccgaagtgcc ggtgacattc gacgcgcggca	36720
cgtaagttag gatgcgggcg ccgtgttggg ccgcgggtcg cgcgacggcc atgaccagcc	36780
gggcgtcggtc gatcaattgc ccgtcgtacg cgagcagacc accgtcgagg ccgtcccc	36840
gaacgggtggg agcaatctcc accaccgtg acggccggat tcggcgcgtat cggggcaacg	36900

tcgcccggc cgtaaccgtc agcacccgca aagcgtcgcc ggccaggaaa cccgcacgca 36960
 ccaacgccccg cttggtgtga cccatcgacg gcaacaacgg gaccagttgc ggcacatggcat 37020
 gcacgagatg aggagcggtt cgtgtcatca ggattcccg cg ttcgacggcg ctgcgcggg 37080
 cgatgcccac gttgccgctg gccagatagc gcagaccgccc gtgcaccaac ttgcagctcc 37140
 agcggctggt gccgaacgccc agatcatgct tttccaccaa ggccaccgccc agaccgcggg 37200
 tggcagcata taaggcaatg ccaacaccgg taatgccccc gcctatcacg atgacgtcga 37260
 gtgcgccacc gtcggccagt gcggtcaggt cggcggagcg acgcgcgcgc ttgagtgcag 37320
 ccgagtgggg catcagcaca aatatccgtt cagtgcgtgg gtaagttcgg tggccagcgc 37380
 ggccgaatcg aggatcgaat cgacgatgtc cgccggactgg atggtcgact gggcgatcag 37440
 caacaccatg gtcgccagtc gacgagcgtc gccggagcgc acactgccc accgctgcgc 37500
 cactgtcagc cggccggcca acccctcgat caggacctgc tggctggcgc cgaggcgctc 37560
 ggtgatgtac accctggcca gtcggcgtt cagtgcgtgg gtaagttcgg tggccagcgc 37620
 caaccggcgtcg gccaccgcga caatctgctt taccaacgt tcccggtcgt ccccggtcag 37680
 gggcacctcc cgcagcacgt cggcgatatg gtcggcgtcgc atggacgcgc tgatcgaccg 37740
 ggtgtccggc cagcgacggt atacggcgtt gtcggcgtcgc cccgcgcgc gggcgatctc 37800
 ggcaagtgtc accccggtcca cgcgttaatc gacgacgcag ctcgcgcgtc cccgcaggat 37860
 acgaccaccg gatccgcgc ggtcattact cattgacagc atgtgtata ctgtaaacgcg 37920
 tgactcaccg cgaggaactc cttccaccga taaaatggc cgcgtgggg gatccgcgc 37980
 cggccaagcc actttctgat ggcgtccggt ctttgctgaa gcagggtgtg ggcctagcgg 38040
 actcggagca gcccgaactc gacccgcgc aggtgcagct ggcggcgtcc gcccgtcgg 38100
 gggcagacca 38110

<210> 25

<211> 2540

<212> DNA

<213> Homo sapiens

<400> 25
 gaaaagggtgg acaagtccata ttttcaagag aagatgactt ttaacagttt tgaaggatct 60
 aaaacttgg tacctgcaga catcaataag gaagaagaat ttgtagaaga gtttaataga 120
 ttaaaaactt ttgctaattt tccaaatggt agtcctgttt cagcatcaac actggcacga 180
 gcagggtttc ttataactgg tgaaggagat accgtgcgtt gcttttagtt tcatgcagct 240
 gtagatagat ggcaatatgg agactcagca gttgaaagac acaggaaagt atccccaaat 300

tgcatgttta tcaacggctt ttatcttgc aaatgtgcca cgcagtctac aaattctggt	360
atccagaatg gtcagtcaca agttgaaaac tatctggaa gcagagatca ttttgcctta	420
gacaggccat ctgagacaca tgcaactat ctttgagaa ctggcaggt tgttagatata	480
tcagacacca tatacccgag gaaccctgcc atgtattgtg aagaagctag attaaagtcc	540
tttcagaact gcccagacta tgctcaccta accccaagag agtttagcaag tgctggactc	600
tactacacag gtattggtga ccaagtgcag tgctttgtt gtggggaaa actgaaaaat	660
tgggaacctt gtgatcgtgc ctggtcagaa cacaggcgc accttcctaa ttgcttcttt	720
gttttgggcc ggaatcttaa tattcgaagt gaatctgatg ctgtgagttc tgataggaat	780
ttccccaaatt caacaaatct tccaagaaat ccatccatgg cagattatga agcacggatc	840
tttacttttgg gacatggat atactcagtt aacaaggagc agcttgcag agctggattt	900
tatgcttttag gtgaaggtga taaagtaaag tgctttcact gtggaggagg gctaactgat	960
tggaagccca gtgaagaccc ttggaaacaa catgctaaat ggtatccagg gtgcaaataat	1020
ctgttagaac agaagggaca agaatatata aacaatattc atttaactca ttcaactttag	1080
gagtgtctgg taagaactac tgagaaaaca ccatcactaa ctagaagaat tgatgatacc	1140
atcttccaaa atcctatggt acaagaagct atacgaatgg ggttcagttt caaggacatt	1200
aagaaaataa tggagggaaa aattcagata tctggagca actataaattc acttgaggtt	1260
ctgggtgcag atctagtgaa tgctcagaaa gacagtatgc aagatgagtc aagtcagact	1320
tcattacaga aagagattag tactgaagag cagctaaggc gcctgcaaga ggagaagctt	1380
tgcaaaatct gtatggatag aaatattgct atcgaaaaat ttccttgg acatctagtc	1440
acttgtaaac aatgtgctga agcagggtac aagtgtccca tgtgctacac agtcattact	1500
ttcaagcaaa aaattttat gtcttaatct aactctatag taggcattttt atgttggttct	1560
tattaccctg attgaatgtg tgatgtgaac tgactttaag taatcaggat tgaattccat	1620
tagcatttgc taccaagtag gaaaaaaaaat gtacatggca gtgttttagt tggcaatata	1680
atctttgaat ttcttgattt ttcagggtat tagctgtatt atccattttt tttactgtta	1740
ttaattgaa accatagact aagaataaga agcatcatac tataactgaa cacaatgtgt	1800
attcatagta tactgattt atttctaagt gtaagtgaat taatcatctg gatTTTTT	1860
tctttcaga taggcttaac aaatggagct ttctgtatataa atgtggag attagagtta	1920
atctccccaa tcacataatt tgttttgtgt gaaaaaggaa taaatttttc catgctggtg	1980
gaaagataga gattttttt agaggttgg tttttttt taggattctg tccattttct	2040
tgtaaaggaa taaacacgga cgtgtgcgaa atatgtttgt aaagtgattt gccatttttg	2100
aaagcgtatt taatgataga atactatcga gccaacatgt actgacatgg aaagatgtca	2160
gagatatgtt aagtgtaaaa tgcaagtggc gggacactat gtatgtctg agccagatca	2220

aagtatgtat gttgttaata tgcataagaac gagagattt gaaagatata caccaaactg	2280
ttaaatgtgg ttctcttcg gggagggggg gattggggga ggggccccag aggggttta	2340
gaggggcctt ttcactttcg actttttca ttttgttctg ttccgattt ttataagtat	2400
gttagaccccg aagggttta tgggaactaa catcagtaac ctaaccccg tgactatcct	2460
tgctcttcc tagggagctg tgggtttcc caccaccac cttccotct gaacaaatgc	2520
ctgagtgctg gggcactttg	2540

<210> 26

<211> 103

<212> RNA

<213> Homo sapiens

<400> 26

agcuccuaaua acaaaaagucu guugcuugug uuucacauuu uggauuuccu aauauaaugu	60
ucucuuuuua gaaaaggugg acaaguccua uuuucaagag aag	103

<210> 27

<211> 28

<212> RNA

<213> Homo sapiens

<400> 27

ggauuuccua auauaauguu cucuuuuu	28
--------------------------------	----

<210> 28

<211> 1619

<212> DNA

<213> Homo sapiens

<400> 28

ccggcagatt tgaatcgccg gacccgttgg cagaggtggc ggccggccgca tgggtgcccc	60
gacgttgccc cctgcctggc agcccttct caaggaccac cgcatctcta cattcaagaa	120
ctggcccttc ttggagggct gcgcctgcac cccggagcgg atggccgagg ctggcttcat	180
ccactgcccc actgagaacg agccagactt ggcccaagtgt ttcttctgct tcaaggagct	240
ggaaggctgg gagccagatg acgaccccat agaggaacat aaaaagcatt cgtccggttg	300

cgcttcctt tctgtcaaga agcagttga agaattaacc cttggtaat tttgaaact 360
 ggacagagaa agagccaaga acaaaaattgc aaaggaaacc aacaataaga agaaagaatt 420
 tgagggaaact gcbaagaaag tgcgccgtgc catcgagcag ctggctgcca tggattgagg 480
 cctctggccg gagctgcctg gtcccagagt ggctgcacca cttccagggt ttattccctg 540
 tgccaccag cttccctgtg ggcccttag caatgtctta ggaaaggaga tcaacatTTT 600
 caaatttagat gtttcaactg tgctcctgtt ttgtcttcaa agtggcacca gaggtgcttc 660
 tgcctgtgca gcgggtgctg ctggtaacag tggctgctc tctctctc tctttttt 720
 gggggctcat ttttgctgtt ttgattccccg ggcttaccag gtgagaagtg agggaggaag 780
 aaggcagtgt ccctttgtct agagctgaca gctttgtcg cgtggcaga gccttccaca 840
 gtgaatgtgt ctggacctca ttttggtag gctgtcacag tcctgagtgt ggacttggca 900
 ggtgcctgtt gaatctgagc tgcagggtcc ttatctgtca cacctgtgcc tcctcagagg 960
 acagttttt ttttggtag tttttttttt ggtagatgca tgacttgtgt 1020
 gtgatgagag aatggagaca gagtccctgg ctccctact gtttacaac atggctttct 1080
 tattttttt gaattgttaa ttcacagaat agcacaaact acaattaaaa ctaagcacaa 1140
 agccattcta agtcattggg gaaacggggt gaacttcagg tggatgagga gacagaatag 1200
 agtgatagga agcgtctggc agataactcct ttgccactg ctgtgtgatt agacaggccc 1260
 agtgagccgc gggcacatg ctggccgctc ctccctcaga aaaaggcagt ggctaaatc 1320
 ctttttaaat gacttggctc gatgctgtgg gggactggct gggctgctgc aggccgtgt 1380
 tctgtcagcc caaccttcac atctgtcacg ttctccacac gggggagaga cgcagtccgc 1440
 ccagggtcccc gttttttttt gaggcagcag ctcccgcagg gctgaagtct ggctaaat 1500
 gatggatttg attcccttc ctccctgtca tagagctgca gggtggattt ttacagcttc 1560
 gctggaaacc tctggaggtc atctcggctg ttccctgagaa ataaaaagcc tgtcatttc 1619

<210> 29

<211> 27

<212> RNA

<213> Homo sapiens

<400> 29

ggcgucacac cuucggguga agucgcc

27

<210> 30

<211> 27

<212> RNA

<213> Homo sapiens

<400> 30

ggcgucacac cuucggguga agucgcc

27

<210> 31

<211> 12

<212> PRT

<213> Homo sapiens

<400> 31

Tyr Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro
1 5 10

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US02/11757

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/68; C07H 21/02; G01N 27/26
 US CL : 435/6; 536/23.1; 204/451

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
 U.S. : 435/6; 536/23.1; 204/451

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
 STN, EAST

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6329146 B1 (Croke et al) 11 December 2001 (11.12.2001), column 40, example 11.	1
Y	US 5,807,682 A (Grossman et al) 15 September 1998 (15.09.1998), column 19, lines 2-18. 61	1
Y	US 6,355,428 (Schroth et al) 12 March 2002 (12.03.2002), column 8, lines 64-67.	1
Y	US 6,320,040 B1 (Cook et al) 20 November 2001 (20.11.2001), column 11, lines 14-22	1
Y	US 6,391,542 B1 (Anderson et al) 12 May 2002 (12.05.2002), column 36, example 18.	1

<input type="checkbox"/>	Further documents are listed in the continuation of Box C.	<input type="checkbox"/>	See patent family annex.
*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier application or patent published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Z"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 22 June 2002 (22.06.2002)	Date of mailing of the international search report 18 JUNE 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703)305-3230	Authorized officer Jyothsna Venkat SRI JYOTHSNA VENKAT PH.D Telephone No. (703) 308-1235 EXAMINER TECHNOLOGY CENTER 1600